

10-30-00

4

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Attorney Docket No. 0635MH-40874

In re Application of:

FREDERICK S.M. HERZ, ET AL.

Serial No. TO BE ASSIGNED

Filed: HERewith

For: SECURE DATA INTERCHANGE

§
§
§
§
§
§
§
§

TRANSMITTAL

Assistant Commissioner for Patents
Washington, D.C.

Sir:

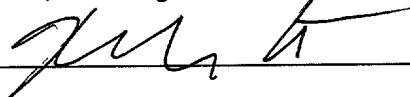
Enclosed for filing please find:

1. This Transmittal with Certificate of Express Mail;
2. Patent Application including Drawings;
3. Declaration;
4. Assertion of Entitlement to Small Entity Status;
5. Power of Attorney;
6. Our check in the amount of \$355.00;
7. Our return postcard which we would appreciate you date stamping and returning to us.

"EXPRESS MAIL" NO. EL643426463US

I hereby certify that this paper or fee is being deposited with the United States Postal Service as Express Mail "Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated below and is addressed to the Commissioner of Patents and Trademarks, Box Patent Application, Washington, D.C. 20231-9998.

By:



Date of Deposit:

27 Oct 2000

10/27/00
jc961 U.S. PTO

jc945 U.S. PTO
09/699098
10/27/00

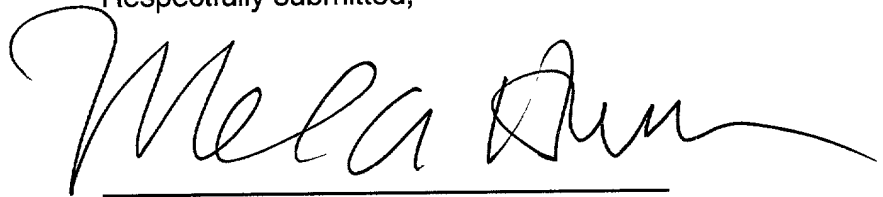
09699098 102700

The filing fee is calculated as follows:

Basic Fee	\$355.00
	<hr/>
Total	\$355.00

No other fee is deemed necessary, however, the undersigned hereby authorizes the Commissioner to charge any additional fees, or credit any overpayments, to Deposit Account No. 50-1060.

Respectfully submitted,



Melvin A. Hunn
Reg. No. 32,574
Kenneth C. Hill
Reg. No. 29,650
HILL & HUNN, LLP
201 Main Street, Suite 1440
Fort Worth, Texas 76102
(817) 332-2113 (voice)
(817) 332-2114 (facsimile)
melhunn@hillandhunn.com (e-mail)
kenhill@hillandhunn.com (e-mail)

ATTORNEY FOR APPLICANT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Attorney Docket No. 0635MH-40874

In re Application of:

FREDERICK S.M. HERZ, ET AL.

Serial No. **TO BE ASSIGNED**

Filed: **HEREWITH**

For: **SECURE DATA INTERCHANGE**

§
§
§
§
§
§
§
§

**ASSERTION OF ENTITLEMENT TO SMALL ENTITY STATUS
UNDER 37 C.F.R. § 1.27(c)**

Assistant Commissioner for Patents
Washington, D.C.

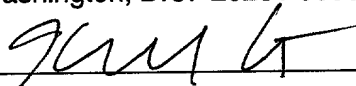
Sir:

Pursuant to 37 C.F.R. 1.27(c)(2)(i), the undersigned hereby asserts that **HERZ TECHNOLOGIES INCORPORATED**, owner by assignment of the entire right, title, and interest in the subject application, is a small entity as defined in 37 C.F.R. § 1.9(d) and is entitled to small entity status for purposes of paying reduced fees under Section 41 (a) and (b) of Title 35, United States Code, to the Patent and Trademark Office with regard to the subject invention.

"EXPRESS MAIL" NO. EL643426463US

I hereby certify that this paper or fee is being deposited with the United States Postal Service as Express Mail "Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated below and is addressed to the Commissioner of Patents and Trademarks, Box Patent Application, Washington, D.C. 20231-9998.

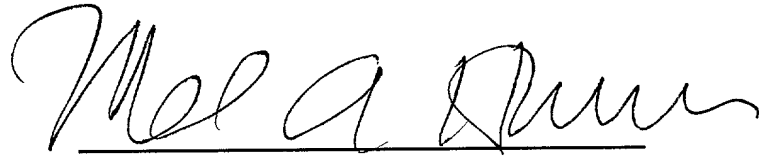
By:



Date of Deposit:

27 Oct 2000

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Melvin A. Hunn". The signature is fluid and cursive, with a horizontal line drawn underneath it.

Melvin A. Hunn
Reg. No. 32,574
Kenneth C. Hill
Reg. No. 29,650
HILL & HUNN LLP
201 Main Street, Suite 1440
Fort Worth, Texas 76102
(817) 332-2113 (voice)
(817) 332-2114 (facsimile)
melhunn@hillandhunn.com (e-mail)
kenhill@hillandhunn.com (e-mail)
ATTORNEY FOR APPLICANT

002207, 26066960

SPECIFICATION

Docket No. 0635MH-40874

TO ALL WHOM IT MAY CONCERN:

BE IT KNOWN that we, Frederick S. M. Herz, David C. Parks, and Sampath Kannan, residing in the state of Pennsylvania, Paul Labys, residing in the state of Utah, and Jason Eisner, residing in the state of New York, have invented new and useful improvements in a

SECURE DATA INTERCHANGE

of which the following is a specification:

CROSS REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of United States Provisional Application No. 60/161,640, filed October 29, 1999, titled SECURE DATA INTERCHANGE, and Provisional Application No. 60/206,538, filed May 23, 1999, titled SECURE DATA INTERCHANGE, both of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention:

The Secure Data Interchange invention describes a system to allow a privacy-protected market for data exchange between multiple self-interested parties. The system presents a general infrastructure for the exchange of information within a safe privacy-protected environment, between multiple self-interested parties. We propose a central data warehouse that maintains data submitted by different users, and executes queries and programs on the data. Rules are associated with data that define how the data can be used and queried, to allow agents that submit data to maintain absolute control over its use. SDI acts as a trusted-intermediary to all parties, and implements an internal market for queries on the information, allowing agents to specify prices for data access. Furthermore, SDI supports complex queries such as collaborative filtering, that can provide a querying agent with a one-time benefit of data access but without long-term access to the data that was used to compute valuable results.

The invention relates to systems that provide personalized information, profiling, automated matchmaking and information exchange, providing a framework that protects privacy and allows information collection and profiling within a carefully controlled environment. Although the marginal cost of data duplication is small, there are hidden costs associated with data, for example because of privacy concerns, and data can be a valuable resource in many problems. In business-to-consumer (B2C) applications Secure Data Interchange addresses the direct conflict between the goal of personalization and the need for privacy, preventing the exchange and collection of information without knowledge and

1 consent. In business-to-business e-commerce applications (B2B) SDI allows vendors to
2 provide sensitive and valuable information, for example about business needs and customer
3 bases, in a secure environment that controls access and leverages value.
4

5 2. Description of the Prior Art:

6 The invention of Secure Data Interchange relates to a wide-range of application
7 domains, all of which are characterized by a need to exchange information in a privacy-
8 protected and carefully controlled market-based environment.

9 As a key application we suggest a system for personalized information delivery in a
10 networked environment, in which the SDI-proxy can fact as a local filter on information,
11 based on what it knows about a user's preferences and methods for filtering pushed by the
12 provider of content. The system allows collaborative filtering through information that is
13 provided to the central data warehouse, but never released directly to other agents;
14 collaborative filtering methods are computed in the central SDI data warehouse. Further
15 motivation is provided with reference to some electronic commerce applications, that we
16 describe in (A) business-to-consumer and (B) business-to-business e-commerce
17 applications.

18 In addition to applications within commerce, the system of Secure Data Interchanges
19 is central to developing many other new products. Examples include the formation of "self-
20 help" groups between a set of individuals with common interests, and applications to
21 personal information delivery systems, e.g. for educational and informational purposes.
22

23 A. Business-to-Consumer (B2C) Electronic Commerce.
24

25 The recent explosion of electronic commerce, in particular Internet-based individual-to-
26 business electronic commerce, presents new opportunities for automated personalized
27 information delivery and the automated customization of products and services. This type of
28 personalization is very valuable to vendors because it can increase sales volumes, enable
29 cross-selling and up-selling of goods and services, and allows vendors to price products

1 dynamically based on information about the preferences and goals of customers.
2 Personalization is also useful to customers when it correctly identifies the requirements and
3 preferences of a customer, because it can reduce search cost and enhance the “shopping
4 experience”. Perhaps a customer can find the good or service (i.e. desirable
5 price/quality/feature tradeoff) that he/she wants more quickly than without personalization,
6 or receive information about an interesting new product or service that he/she did not know
7 about.

8 The basis for these new services is that Internet-based “shop fronts” can be
9 individualized on a per-customer basis, dynamically and in real-time. Traditional main-street
10 shops must offer the same store layout to every customer, because the layout is physical,
11 although some level of personalized service can be achieved through well-trained sales
12 assistants, that act as a “guide” for a customer within a store. On-line “shop fronts” are
13 virtual, and configurable at negligible cost to the customer or the vendor, assuming that
14 computation is cheap and fast.

15 Furthermore, Internet-based electronic commerce can allow business to collect vast
16 amounts of consumer information, because customers interact through a computer-based
17 interface. Customers can be monitored as they browse a Web site for products and services.
18 Information such as the search-terms that users enter into a search engine, the links that
19 users follow, and the length of time spent on each page, can all provide an insight into the
20 current goal of a customer, i.e. the type of product that he/she wants. When combined across
21 different sessions, and with similar information about the browsing and purchasing habits of
22 other customers, the information can be folded into a long-term view of the preferences and
23 needs of a customer.

24 Moreover, new network connectivity enables different vendors to exchange profiles
25 for common customers, either statically or dynamically, in order to build broad and detailed
26 profiles across vendor domains. There exist many potentially powerful synergies between
27 the data sets that are collected by different vendors, that can be leveraged to provide
28 appropriate services and products to customers. When analyzed with the proper statistical
29 tools these data sets can reveal fundamental patterns in the behavior of users, and enable a
30 vendor to provide appropriate information to a user. Furthermore, access to user-profiles

collected by other vendors can enable vendors to provide focused information delivery to first-time users, and also cross-market services with other appropriate vendors.

Providing user profile information within a carefully controlled environment can benefit vendors and users:

- Vendors would find benefit in sharing data with other vendors; this would deepen their understanding of their customers' behaviors and preferences, especially if some customers were traceable across several data sets.
- Users would benefit from sharing data with other users. This is already evident in the popularity of news groups and web discussion pages catering to individuals with shared interests. By learning what other people with similar tastes and preferences have discovered and enjoyed, a user can sidestep information overload in the search for personally satisfying information.
- Vendors can benefit from receiving data about users. An obvious example would be in the use of collaborative filtering for the marketing of targeted promotions; rather than being deluged with coupons and advertisements that are of absolutely no interest, a user would benefit by being presented with advertising that is highly relevant. In the process, the vendor would increase advertising response rates, boosting overall efficiency.
- Users can receive benefits from providing information to vendors. Personalization of content at vendors' web pages, and well-focused banner advertisements at other web sites that they visit.

The problem is that a user wants controlled personalization, in the sense that it might not be desirable for information about every on-line transaction that a user performs, every on-line document that a user reads, and every web page that a user visits, and demographic information, to be available to every business that the user interacts with, in the virtual and physical world.

A.1 Focused Banner Advertising/Content provision

Internet-based media sites have followed preceding formats in generating revenue from advertising, with content to users often provided free-of-charge. The business model is

1 similar to that in newspapers, magazines, and television, where circulation and
2 audience/readership demographics are used to drive revenue. Electronic media presents new
3 opportunities for media-based business: for example multimedia techniques and
4 interactivity, personalized delivery of information, and personalized targeting of
5 advertising.

6 The problem – as before, is to acquire and leverage information about the
7 preferences and interests of a user, within a system that protects user privacy (i.e. controls
8 the collection and exchange of information about users, and controls the use that is made of
9 that information). A further problem is to extrapolate information from a large corpus of
10 data about an individual user.

11

12 **A.2 Mailing Lists**

13

14 As another example, suppose that business A requests a list of individuals that meet a
15 particular criteria. Consumer B meets the criteria, but is only listed for business A if A also
16 meets criteria specified by B, for example if A will provide information about new products
17 and services that are interesting to B. In an application to the profiling of users on-line, the
18 problem is that users want to receive the benefits of targeted products and advertisements,
19 but want to avoid the abuse of profile information and control vendors' access to that
20 information.

21

22 **B. Business-to-Business (B2B) Electronic Commerce**

23

24 The Internet provides businesses with network connectivity with other business, both
25 competitors and partners. This connectivity allows businesses to exchange information
26 about customers (dynamically or statically), in order to identify potential new customers,
27 build better profiles for existing customers, and up-sell/cross-sell products and services in
28 real-time. The problem with this exchange of information (that can include swaps, sells, and
29 rental access) is that businesses need to (a) protect the privacy of their customers; (b)
30 prevent information release to competitors, either directly or through third-parties.

B.1 Privacy-Protected Identification of Synergies/Matches

There are many scenarios where autonomous agents would like to be informed of matches under conditions of mutual consent, but without information leakage to any agent if any one of the agents declines the match. Consider two vendors, A and B, and suppose the vendors seek strategic partnerships with other vendors that have appropriate skills and goals. However, vendor A does not want to broadcast to all vendors its need for a business partner or a new alliance, instead vendor A wants to be introduced to another vendor with the right mix of capabilities; similarly for vendor B. What is required is a system that only introduces vendor A to vendor B, and perhaps anonymously at first, if both vendors consent to the introduction. The problem is to provide information that enables matches, without allowing bad matches and abuse of information – i.e. within an environment of secure data interchange.

B.2 Credential-based Introductions, Contracting and Messaging-systems.

There are many situations where individual parties, for example individuals or businesses, require introductions to credentialed individuals and/or businesses, with the aim of building a new relationship or making a new contract. Consider for example business associations, where credentials about non-bankruptcy, and no previous attempts to defraud could be important. Consider social introductions, where individuals might be concerned about past criminal activities of new contacts. In the domain of automobiles, we could consider a system that identifies other automobiles in the physical location of a vehicle that have recently been involved in an accident. The problem is to manage certificates within a system where users can maintain multiple identities, and to protect the release of certificates without suitable provisions for terms-of-use and criteria for request.

SUMMARY OF THE INVENTION

1 The above problems are solved, and a technical advance achieved, by the system of
2 Secure Data Interchange. The Secure Data Interchange system enables information about
3 bilateral and multilateral interactions between multiple persistent parties to be exchanged
4 and leveraged within an environment that uses a combination of techniques to control access
5 to information, release of information, and matching of information back to parties.

6 The system of Secure Data Interchange (SDI) provides a trusted server containing
7 a large database of information that is owned by its providers. Each data record has an
8 associated price rule, that controls access to data. The pricing model allows a data owner
9 to specify a price for different types and amounts of information access, and whether the
10 identity of the information owner is required, and the system of SDI computes a
11 composite price for a query based on aggregated prices for a query over a number of
12 different data owners, with an internal market that favors low priced data. The pricing
13 model allows discounts based on certificates of a requesting agent, and as a special case
14 implements the standard capability-based access control systems, where information is
15 provided to users with appropriate permissions (i.e. with zero and infinite prices). In
16 addition, the system of Secure Data Interchange allows data to be submitted with a level
17 of random perturbation (noise), to provide added privacy protection, or alternatively
18 allow an agent to specify in conditions under which additional noise should be added to
19 data. A query is priced before execution, to allow an agent to decide whether or not to
20 execute a query, and select between alternative types of queries. Binding price quotes are
21 provided to querying agents, and queries can be scaled to meet a budget.

22 Data owners can submit data to the central data warehouse with different degrees
23 of identification, for example anonymously, pseudonymously, or with a true identity. For
24 example, in the case of data that represents a user's profile information, for example
25 information about the interests of a user, a user might prefer to use a number of different
26 pseudonyms for different types of activities that he/she likes to engage in online. A user
27 might maintain a number of different aliases within the database, for example to represent
28 different types of things he/she likes to do which have little bearing on each other.

1 Various types of queries can be executed on the server, ranging from traditional
2 SQL style queries to collaborative-filtering style queries. The inventions of SDI is not
3 predicated on the type of queries supported, describing instead a general system to
4 execute those queries within a secure data-controlled environment. We leave the
5 algorithms that implements the semantics of a query undefined, but define the interface
6 between that algorithm and the data in the database, which is protected with price rules.

7 A query can perform a considerable amount of computation on records in the
8 database before any information is provided in response to a query. We allow general
9 purpose programs to run on the server, that might for example perform collaborative
10 filtering or other data mining techniques before returning aggregate information, for
11 example a new model of the profiles of users of a particular interest to a vendor.
12 However, the total charge for a complete query session is computed as the sum payment
13 charged by all record pointers that are used in computing the final response.

14 Furthermore, queries can also take an action on behalf of a querying agent, for
15 example asking another agent to provide more information about something. We allow
16 queries to be priced, based on a model of "revenue collection", in which the SDI server
17 sells the right to access to data on behalf of the agents that submit data into the database.
18 We also propose to allow "persistent queries", which reside on the server for a fixed
19 period of time and return a response to the querying agent whenever conditions are met in
20 the data.

21 Interesting variations of SDI place data in different distributed locations, and
22 move the control of information access between a central SDI server and distributed
23 client-side SDI proxy agents. This allows different tradeoffs between privacy and
24 information sharing. It also has implications for bandwidth and computational
25 requirements within SDI. One role of a client-side SDI data warehouse is to provide the
26 same functionality as the central shared SDI database, but with processing only
27 performed on information provided by that agent. This can allow greater privacy by
28 allowing a user to retain absolute control over his/her data on his/her local machine
29 without even releasing data to the shared database.

30 In an application to personalized on-line interactions, we describe a client-side
31 SDI proxy which manages a user's interactions with the on-line sites of vendors and also

1 manages a user's interactions with the central SDI data warehouse, i.e. providing profile
2 information and controlling profile access. The client-side SDI proxy for an agent that
3 represents an individual browsing the Internet can manage that user's profiles in
4 interactions with other agents, for example representing vendors and content providers.
5 The client-side SDI proxy can also handle decisions about what types of information to
6 submit to the server, and manages query execution on behalf of the agent. The client-side
7 SDI proxy agent can also push information about a user's on-line activities to the central
8 SDI data-warehouse in real time. This enables a system of "time-of-purchase-
9 competition" system, in which a user can request competitive counteroffers from other
10 vendors before making a purchase.

11 The system addresses the fundamental conflict that exists between rights of privacy
12 and efficiency gains from better bilateral exchange of profile/preference information. SDI as
13 applied to B2C e-commerce allows consumers to receive targeted information about
14 products and services, but without the loss-of-privacy that can easily occur in the current on-
15 line profiling "free-for-all". The cookie technology provided by Netscape to supported
16 personalized sessions with a single vendor on-line has been used by advertising network
17 providers such as DoubleClick to track users across multiple sites, often without either the
18 consent or knowledge of that individual [New York Times, Feb 7, 2000].

19 In describing the system of secure data interchange we claim the following novel
20 technical ideas:

- 21 (a) Agents can associate price-rules with information that is placed in the central
22 database, and retain absolute control and ownership over all uses of that data.
23 As a special case of price rules, the system supports access based on certified
24 properties of querying agents (with zero and infinite prices). The pricing
25 model allows information providing agents to receive direct value for data,
26 and allows agents that request access to information to receive a price before
27 a query is executed, and make appropriate decisions about what type of
28 queries to execute.
- 29 (b) A number of novel techniques are proposed to allow data processing within
30 the data warehouse without releasing too much information to an agent.
31 General programs, for example collaborative filtering techniques, can be

1 executed in situ within the data warehouse, so that agents can receive the
2 aggregate benefits of information, without receiving details about the
3 information. We might also allow an agent to receive anonymous or
4 randomized information, and control the amount of information that can be
5 received from a particular record.

6 (c) The system of SDI can act as a trusted intermediary between agents,
7 notifying agents about information of a particular type, and for example
8 sending messages between agents without breaking the identity of agents
9 except by consent. A special type of query that we call a persistent query
10 allows an agent to maintain a permanent “searching” presence in the central
11 database, always on the lookout for useful information.

12 (d) In a key variation we push a small version of the SDI data warehouse onto
13 an user’s client computer, where it acts as a proxy agent, and further proxies
14 an agent’s interaction with other agents in real time. The local SDI data
15 warehouse, called the client-side SDI database, is a trusted party to the
16 client, trusted to maintain information that it is valuable to the client and use
17 that information in appropriate ways. The advantage of this method is that an
18 individual never needs to release sensitive profile information, it is always
19 held on its local computer—but can still receive the benefits of
20 personalization.

21 (e) We also suggest a client-side SDI proxy that can collect information about a
22 user, for example within an Internet browsing application, and periodically
23 push the collected information to the SDI data warehouse in a controlled
24 way. The client-side SDI proxy can also be responsible for certain data
25 certification functions, and can manage a user’s interactions with other
26 agents to protect its privacy in non-SDI mediated transactions.

27 As an application to B2C e-commerce, the system of SDI allows client-side
28 personalization instead of provider-side personalization. Instead of passing profile
29 information to a provider and receiving personalized information in return, providers can
30 provide personalization methods that are used interactively with local profile information
31 about consumers to target products and services without receiving explicit information

BRIEF DESCRIPTION OF THE DRAWINGS

1 The novel features believed characteristic of the invention are set forth in the
2 appended claims. The invention itself however, as well as a preferred mode of use, further
3 objects and advantages thereof, will best be understood by reference to the following
4 detailed description of an illustrative embodiment when read in conjunction with the
5 accompanying drawings, wherein:

6 Figures 1 through 21 illustrate various parts and embodiments of the invention.

002201" 85066960

DESCRIPTION OF THE PREFERRED EMBODIMENT

1. Introduction

The invention of Secure Data Interchange (SDI) describes a general infrastructure for the exchange of information within a controlled environment. We propose a central data warehouse that maintains data submitted by different users, and executes queries and programs on the data. Rules are associated with data that define how the data can be used and queried. As such, the system of SDI prevents the exchange and collection of information without knowledge and consent. The system allows for payments to be received by the providers of information, in return for data access. In application, the invention enables new systems for the delivery of personalized information, profiling and automated matchmaking and information exchange, all within a framework that protects privacy and maintains data security. SDI supports the collection and exchange of information between, and relating to, autonomous (and possibly self-interested) agents within a distributed environment.

1.1 Definitions.

Agent. An agent in SDI is a any party that wants to sell or give away data to other parties, or buy or receive data, or in general both provide and receive data. Agents may represent any party with individual goals, autonomy of control, and a persistent identity. Examples in business-to-consumer e-commerce include business such as newspapers, book stores and travel companies, that wish to receive data about the profiles and buying habits of users so that they can personalize the information, products and services that are sold to users. We assume that agents are autonomous from the system of Secure Data Interchange, and follow actions consistent with their preferences, abilities and resources. Similarly, an on-line consumer is an agent that wishes to provide data about its preferences and buying habits to vendors and other consumers, in return for well targeted products and financial reward.

Agent Computers. Agents are represented in the system of secure data interchange with dedicated computational resources, agent computers, with permanent memory, processing power, and network connectivity. For example, an agent computer might be a consumer's home PC that will act as a client machine in interactions with the central SDI data server and

the servers of on-line vendors. For a vendor, an agent computer might be the server computer that it uses to execute its on-line business. Agent computers might also be thin clients, such as mobile computing devices, handheld devices, cell phones. We push different amounts of data and functionality within SDI to agent computers and away from web centric devices depending on the nature of the computer and an agent's preferences.

Certificates. Certificates are used within SDI to establish trust between different agents, and to help agents to reach useful agreements.

Profile: Each agent can have one or more profiles, which the agent can assume in its interactions with other agents. Part of a profile is a user's **identity**, that provides a (possibly limited) method for another agent to identify the agent in the future when it assumes the same profile. We allow three types of agent identities: anonymous, pseudonymous, persistent pseudonymous, and true identity. An agent may assume a profile and an identity in any interaction with another agent.

- Anonymous. To assume an anonymous identity agent A creates a one-time identifier that it uses in interactions with one other agent. The identifier may allow the other agent to respond zero or one times.
- Pseudonymous. A pseudonymous identifier created by agent A may be used with more than one other agent, and allows agents to respond to agent A as many times as they like, but agent A can terminate the pseudonym at any time and separate from the identity. Agent A can optionally restrict the number of agents that may reply to the pseudonym, for example to the agents that it explicitly provides with the identifier.
- Persistent pseudonymous. A persistent pseudonymous identifier is created by agent A for use with agents in set S, and provides the added condition that agent A promises to use the same pseudonym for all interactions with all agents in set S for all time into the future. Agent A can optionally restrict the number of agents that may reply to the pseudonym, for example to the agents that it explicitly provides with the identifier.
- True identity. If agent A interacts with another agent under its true identity then it has no method to prevent the agent or any other agent responding to agent A in the future.

1.2 System Architecture

The basic architecture for SDI is a system of agent computers, connected via a network (Internet, wireless, or otherwise) to other agent computers, and with a central SDI shared data warehouse. In Figure 1 we illustrate the top-level architecture of Secure Data Interchange, the networked system of agent computers and a centralized server computer that acts as a repository for data, rules and code. This is called the SDI data warehouse. Each user is associated with an agent computer, and in general users can be individuals, groups of individuals, or companies. In its most general form, the system of Secure Data Interchange is for a system of multiple autonomous agent computers, involved in multilateral communication. We restrict the system description to bilateral communication between agents, without loss of generality because any multilateral (multicast or broadcast) can be implemented as a set of bilateral communications. In any bilateral communication there are two parties, the sender and the receiver (and the parties can dynamically change over a communication session).

SDI proxy agents and profile management policies on agent computers mediate bilateral agent interactions. Agents define profile-management policies that are implemented on agent computers and determine appropriate profiles and identities for an agent in interactions. Every time an agent initiates a new interaction, new information is available about the agent, that can be compiled by the agent's own computer device, and also by the computer device of the agent with which it interacts. Careful profile and identity management provides an agent with absolute control over the ability of other agents to profile the agent and exchange information about the agent, for example the agent can use a unique pseudonymous identifier with every other agent. However, the core of the SDI invention is that we encourage agents to exchange information, by providing a secure central data interchange for that purpose. Agents can submit profile information, and other data, to the central data warehouse and make it available for particular types of data mining by other agents, and receive financial reward for providing data. Products and services can also be targeted for agent A on that agent's computer, without the provider agent receiving information about agent A's profile. For example, vendor B might provide generic information about its products to agent A, and a personalization rule that it has constructed

1 from data mining the central SDI data warehouse. The agent computer of agent A can then
2 use its local and private profile information to decide what products to target to the agent,
3 based on the instructions that it receives from agent B.

4 The arrows indicate possible flows of information between the different elements
5 within the system. We allow agents to communicate directly, via bilateral or multilateral
6 communication, and also indirectly via the central data warehouse. Information can be
7 submitted to the central data warehouse, for example new data and new rules, or queries
8 that the data warehouse will execute. The data warehouse returns information to agents, for
9 example the results of a query. The Secure Data Interchange invention is independent of the
10 implementation details of the communication platform. We assume that the system of SDI is
11 built on top of a secure communications platform, for example via a SSL-encrypted TCP/IP
12 session in an application to the Internet. Furthermore, we draw on cryptographic techniques
13 known in the art for identity management, and additional techniques to support
14 pseudonymous identities within a particular network protocol, e.g. the stripping of sender
15 URL addresses from HTTP packets on the Internet.

16 In overview, the top-level claim in the system of Secure Data Interchange, of a
17 privacy-protected market-based system for data exchange between self-interested parties, is
18 constructed from the following core components:

- 19 ○ A Central Data Warehouse, with associated price-rules and constraints submitted by
20 the owners of data. Architectural variations allow virtual links to data in the central
21 data warehouse, with data physically located in distributed locations.
- 22 ○ A Query-execution and Price module, which executes queries on the data within
23 constraints on the types of queries that can be performed, and computes the price of
24 queries, collecting payment.

25 and the following optional components:

- 26 ○ Distributed data management, via client-side SDI “proxy” agents, that intermediate
27 the interactions between agents, and manage data provided to the central SDI data
28 warehouse. Distributed data management includes methods to manage an agent’s
29 identity in interactions with other agents, for example via pseudonymous and
30 anonymous interactions.

- Distributed query-execution, via client-side SDI query-execution modules, that allow data that is physically stored in distributed locations to be queried on distributed nodes, without an explicit release of the data.
- Data perturbation methods to augment data values with randomized noise, to allow queries to tradeoff price and quality, and to allow a user to protect his/her identity by hiding revealing details of submitted data elements.

In describing key applications of the system of Secure Data Interchange, we describe in later sections additional features that extend the functionality of the invention:

- Community dollars: an extended payment scheme that allows extended forms of payment within the system, for example allowing payment to be made in terms of discounts in purchases from particular vendors, or in return for agreements to purchase a certain number of products over an extended period of time.
- Within an Internet-browser based system, methods to control a user's browsing experience via a personalized portal, where personalization is performed via data mining techniques executed by the system of SDI on data in the central data warehouse.
- Methods to support anonymous and pseudonymous electronic commerce, e.g. delivery and payment services.

In the next section we provide an overview of each of the central SDI data warehouse, which forms the core component of Secure Data Interchange. We then add technical details for specific components, to make our ideas more concrete.

1.3 Implementation Details

In this we describe some of the core technologies known in the art that would be used to build up an implementation of the SDI system. The technologies span areas in cryptography, for pseudonym management, digital certificates, payment mechanisms, etc.; and information theoretic-methods, for example to protect the identity of the originator of a message by routing messages through local "crowds" of agents.

1.3.1 Certificate Management

1 Cryptographic techniques well-known in the art [Chaum 81; Chaum 85; Chaum 91] provide
2 the ability for a certificate to be linked to an agent's identity, and not transferred to other
3 agents. This is the basic functionality required of a certificate management system. For
4 example, a certifying agency can sign the public key of an agent with the private key of the
5 certifying agency, to indicate that the agent satisfies requirements for certification. Another
6 agent can verify the certificate with the public key of the agent and the public key of the
7 certifying agency. The certificate cannot be transferred to another agent unless that agent
8 assumes the same public key. We assume a public key infrastructure to manage this process.

9 **1.3.2 Support for Anonymous and Pseudonymous Identities**

10

11 We have already noted that it is important to provide whatever additional support is required
12 because of the underlying communication infrastructure to protect agents' profile
13 management policies. For example, in the TCP/IP mechanism a message must be stripped of
14 the network address of the originating Internet server, because this can provide information
15 to allow pseudonyms to be linked. Similarly, messages can be routed through a common
16 gateway or random "forwarders" as in the "CROWDS" system [GGMM98; RR98] to
17 provide pseudonymity. Furthermore, other e-commerce functions, such as payment and the
18 anonymous mailing of goods must be supported (see [INSERT A FORWARD REF TO A
19 LATE SECTION OF BOTTOM-LEVEL SDI DOC]).

20 The ability to embed data within web pages allows client-side processing of
21 information. By embedding profile and location information directly within a web document
22 we can alleviate the bandwidth and computational bottlenecks that can occur at a centralized
23 profile server if profiles are fetched on-the-fly when web pages are downloaded by clients.
24 The origin server (supported by the vendor) requests periodic profile updates from the
25 central SDI server. This duplication of information enables the profile and the page contents
26 to be provided directly from a vendor's server.

27 There are some potential drawbacks of this approach: (1) the profile information
28 associated with a web page and target objects can be out-of-date; (2) the profile information
29 is available to all clients and proxy servers, not just those that are SDI-enabled; (3) the
30 profile information can be altered. We suggest technical solutions to each of these problems
31 below.

1 In addition, we prevent unauthorized access of embedded profile information
2 through the encryption of the metadata that is represented within the XML structure of a
3 web page.

4 Profile information can be encrypted using a hierarchy of keys, so that different
5 levels of access to the information may be provided according to the access levels of users
6 and vendors. All users that request web pages from SDI-enabled vendors, whether or not the
7 user is a member of SDI receive the same profile information. We provide encrypted
8 profiles to vendors in the 'Profile Update' messages from SDI to vendor servers, so that: (a)
9 unauthorized agents cannot tamper with the profiles; (b) the profiles cannot be read by
10 unauthorized agents.

11 The SDI system supplies a private key to trusted SDI client software, that enables
12 only SDI-enabled clients to access profile information, and only access that information to
13 the extent permitted by privacy policies of users and vendors. Different levels of encryption
14 enforce multiple levels of access. Periodically the key pairs are changed to prevent
15 extended attempts at cryptographic attacks. The SDI system uploads the key that provides
16 the correct level of access for a user to a user's client, once terms of access and profile
17 management have been agreed. A client can only access embedded information once
18 enabled with a relevant key. Finally, profile information is signed with a digital certificate,
19 to prevent third parties from tampering with profiles for commercial gain.

20 **2. Core Modules**

21
22 Figure 2 illustrates the core modules within the system of Secure Data Interchange. In
23 this section we briefly describe each module in relation to the other modules, and provide
24 more details in later sections.

25 The central SDI data warehouse is the core system in SDI, managing data records on
26 behalf of agents. In combination with the query-execution module, these two modules
27 implement the privacy-protected market for query-execution, where queries are executed:

- 28 • If the querying agent has the right certificates, as defined by the owner of data
- 29 • (and) If the querying agent pays the cost of performing a query

1 Furthermore, the system of SDI implements an internal market, with queries executed
2 as cheaply as possible, within quality constraints provided by a querying agent. The
3 eBank module in SDI collects payments, and allows those payments to be transferred to
4 external bank accounts, or used in part payment for products and services to SDI-enabled
5 vendors.

6 We allow the central SDI data warehouse to be replicated and/or partitioned across a
7 network, for example a small single-user SDI data warehouse can be situated at a user's
8 client machine to allow vendors to provide personalized products and services without
9 releasing personal information (e.g. profiles). The rules that personalize products are
10 queries, which can be executed in the SDI proxy on a user's machine.

11 The central data warehouse acts as a repository for data submitted by agent computers
12 on behalf of users. Consider, for example, data that represents user profiles (i.e. what
13 types of books they like to read, what their political interests are, etc.), or data that
14 represents business rules (i.e. what type of customer a business is seeking, what types of
15 products it offers, what type of strategic agreements it is looking to make with other
16 businesses, etc.)

17 The data repository contains information submitted by agents to the central data
18 warehouse. It is not necessary that a single agent links all information submitted by one
19 agent, because agents can submit information anonymously or pseudonymously.
20 Although agents submit data to the central shared warehouse they maintain key aspects of
21 ownership of that data, but without minute-by-minute management of the data. In
22 particular, providers of information provide the central SDI warehouse with a certain
23 amount of autonomy to control how the information is used: the types of uses that are
24 permitted, the type that are not prohibited, and what types of rewards (financial or
25 otherwise) are required for access to the information. The system of SDI, in the access
26 that it allows to information in the database, acts as a trusted agent for information. One
27 role of SDI is to protect the identity of an agent that requests and provides information—
28 allowing agents to remain anonymous when that is desirable. We outline in this
29 description the key technical solutions with which this is achieved.

30 The data is heterogeneous, best represented within an object-oriented database
31 system. Each data object is associated with an accessor function, which describes the

1 syntax of valid queries on a particular type of data, and executes queries. As data types
2 become standardized, we could suppose a library of standard accessor functions for
3 different data types, perhaps provided by third parties. For example, a special data type
4 could be developed to represent the profiles of online customers. As part of query
5 execution therefore, is the execution of methods associated with heterogeneous types of
6 data. Of course, we can also describe a simple special case where the data types are more
7 transparent to SDI, and data fields are directly indexed as in standard databases.

8 Agents can submit queries to the SDI data warehouse, these are placed in the pending
9 queries queue if they are one-time and to be executed as quickly as possible, or in the
10 persistent queries queue if they are to be executed periodically, as the data in the
11 repository changes. The query execution module contains the processor that performs
12 instructions on behalf of agents, within the constraints of price rules.

13 Price rules are associated with data records. Price rules compute a charge for
14 requested information about the data record. We describe a number of simple price
15 languages, to allow an owner of information to configure access to the data. In general
16 the price rules can be quite complex, and implemented within an object-oriented
17 framework, so that the system of SDI is expandable to new data types. The query-
18 execution module can execute both the methods to access data in objects, and the
19 methods to price access to data. In performing query execution, we suggest some
20 techniques to minimize the price paid to perform a query—for example we implement a
21 simple internal marketplace where the data records with the smallest price are selected,
22 with all other things equal.

23 As a special case, prices can implement the standard capability-based access control
24 systems, such as those in UNIX, where a user must possess the necessary properties to
25 read information. A price \$0 corresponds to permission to access information, while a
26 price \$infinity prevents a user from reading information. The price rule computes a price
27 for accessing information based on a number of factors: the properties of the requesting
28 agent, as demonstrated with certificates; the information requested; and whether the
29 identity of the owner of the information is revealed to the requesting agent. SDI enforces
30 the price rules, with payment collected from agents before the results of queries are
31 reported. SDI allows an agent to contribute information to the shared database but

0969098-102700

1 maintain control over access to the data. Price rules can also specify different prices
2 based on the degree of random perturbation that is performed on data before a query is
3 executed. This can protect the identity of the owner of information but still allow
4 valuable information to be provided in response to a query.

5 The system of SDI also expands the possible space of agreements through anonymity
6 techniques and through random perturbation of data. Agents can provide information
7 anonymously or pseudonymously, or with their true identities revealed. Furthermore,
8 agents can charge less for information provided anonymously than for information
9 provided under a revealed identity. Meta-information associated with data can also
10 specify whether the information has been certified by a third-party, or whether it has been
11 randomized slightly before placing in the SDI data warehouse.

12 The data repository is linked to the query-execution module, that performs queries
13 that are queued to be executed in the pending queries queue. We allow queries to be
14 general programs, an extension from the simple queries that are found in SQL-based
15 database languages. The key novelty is that the query-execution module can maintain
16 intermediate results, for example pointers to records selected as part of intermediate
17 queries. In a standard database there is more of a separation between data access and data
18 processing, with data accessed, pulled outside of the database, and then processed. We
19 allow data to be accessed and processed in the database, with results pushed to clients.
20 This is has useful privacy properties, because less information is finally provided in
21 response to a query. It is very useful in the space of secure data interchange because it
22 expands the set of agreements that can be reached between providers of information and
23 requestors of information.

24 Data manipulation is via record pointers, which are one-time and anonymous pointers
25 to data records. We associate a single query (possibly a complex query, i.e. a query
26 program) with a query session. A record pointer is a temporary identifier for a record that
27 is selected. The pointer allows persistent queries to be performed on the same record
28 during a single query session, but is not valid in other sessions. The same data record
29 might be accessed multiple times during a single session, and with multiple record
30 pointers when the accesses are independent (i.e. when the querying agent does not know
31 that it is requesting information of the same data record). Provisional payments are tallied

1 against each record pointer, and relate to the degree of information provided about that
2 record. For example, consider the problem of finding a set of profiles of a particular type.
3 As the search is performed over data records in the database the query-execution module
4 assesses a provisional payment, as records are accessed. However, the requestor of the
5 information only pays the sum payment charged by the records that are actually selected
6 as suitable. The owners of the data records that were not selected in the final result
7 returned to the user receive no payment, because no information about their data was
8 released and the information was only used in intermediate processing steps. This is
9 important, because it makes it unnecessary to formulate queries carefully in order to
10 avoid extra cost because of redundant intermediate steps.

11 Complex queries, including query-programs, with intermediate results and
12 anonymous record pointers expand the space of data manipulation in a system with
13 providers of information that are more willing to release information if that information is
14 never made directly available to another agent except in aggregated system-wide terms.
15 An agent might be willing to allow submitted data to be used for data mining applications
16 so long as the data remains in the trusted data repository. The central SDI server can
17 provide standard types of query-programs, stored in the SDI-methods module, for
18 example methods to perform collaborative filtering.

19 The query-execution module also performs persistent queries that are submitted by
20 agents to be executed when particular conditions in the data repository are true. Persistent
21 queries check for certain information to reside in the data warehouse, and notify a
22 requesting agent whenever these conditions exist, perhaps automatically making an
23 action.

24 The pricing module is responsible for pricing queries before execution. The basic
25 problem is to aggregate the charge over all data records that provide information to form
26 part of the result of the query, and to estimate the price of a complex query without
27 performing the query. We suggest a simple top-level query language to allow a user to
28 price quality-cost tradeoffs in the query that is finally executed. For example, statistical
29 techniques can be used to compute aggregate statistics without accessing all data records
30 in the database. To give another example, a query can be performed on the basis of
31 accurate information or on the basis of slightly randomized information. The pricing

05699098-102700

1 module reports a price to a requesting agent, and then allocates payment to appropriate
2 agents in the e-Bank if and when a query is actually executed for a user.

3 The e-Bank is a module that maintains a balance of payments received for each data
4 entry in the warehouse, one balance for each of the unique SDI identifiers with which an
5 agent submits information.

6 There can be multiple accounts for a user in the case that the same person or company
7 submits information under multiple IDs (see below).

8 The communication module receives messages from agents. Agent messages can be
9 updates about information or access rules in the data warehouse, information received
10 from the owner of data. Messages might also be new queries. The module also sends
11 messages to agents, for example messages that indicate the results of an instruction are
12 available, or to send messages that are generated by instructions from other agents and
13 consistent with the rules associated with an agent's profile. The module also chooses an
14 action to perform based on messages received, i.e. processes incoming messages and
15 either discards them, places them in the appropriate location if they are a query, or
16 updates the data and rules when a message contains new data.

17 The following sections describe each of the key components in turn.

18 **3. Data Repository**

19 The data repository in the SDI data warehouse is a general-purpose database that stores
20 information submitted by agents. Multiple agents submit information, each of which
21 retains control over access to their data. Access to information is controlled via the price
22 rules that an agent specifies when providing information. The SDI query execution
23 module enforces price rules, collecting revenue from requesting agents before providing
24 results. The system of SDI allows users to receive economic benefits from information,
25 and realize synergies—in a controlled environment. SDI implements an *internal market*
26 *for information*, such that data records with the best price are used in queries, given a
27 number of records that satisfy other requirements.

28 For example: SDI can be used to store the profiles on online customers, to allow
29 vendors to better personalize services and identify market opportunities *and* to allow
30 customers to receive value for their profiles; SDI can be used to store job offers and job

1 applications, where job offers can contain salaries that are only revealed to applicants
2 with suitable qualifications.

3 SDI is designed to allow users to submit many different types of information, from
4 many different sources, and for many different purposes. The application is particularly
5 useful when it becomes a *de facto* data repository for lots of different types of
6 information, information that can in fact be *analyzed* for the purpose of extracting
7 patterns and other useful data (e.g. with collaborative filtering techniques). However, we
8 provide three key variations in the following description, which are not necessarily
9 exclusive.

- 10 1. A single unified database, with information residing on a central server (or a
11 number of central servers). We describe two possible implementations for such a
12 heterogeneous database, one XML-based and another object-oriented model.
- 13 2. Multiple databases, partitioned into "data types". For example, one SDI database
14 might contain profiles on online customers, while another database might contain
15 information about job offers and job applications. This type of database can be
16 implemented using a standard "indexed-field" representation. Again, the
17 information resides on a central server (or a number of central servers).
- 18 3. Distributed data, with some data residing on a central server, and some data
19 physically located on distributed servers (e.g. on the servers of vendors that
20 subscribe to SDI), but with virtual "hyper links" from the central server to give
21 one unified view of the data.

22 The key over-riding problem is one of *data representation*, the way in which information
23 is encoded. We describe a general infrastructure for information exchange, and do not
24 mean to limit the description to any one type of data. Furthermore, the system of SDI
25 should be *expandable*, so that new data types can be introduced. In this specification the
26 type of information in the Secure Data Interchange shared data warehouse is not
27 constrained in any way, but may include for example user profiles (e.g. preferences,
28 recent purchases, etc.) or business services (e.g. costs for services, service capabilities,
29 etc.).

30 One approach is to use a single native SDI ontology, which is expanded as
31 necessary. It would be the responsibility of providers of information in alternative forms

1 to provide "translation services" to convert local data formats into the SDI native data
2 format. A typical technology to support a shared ontological representation is XML
3 (Extended Meta Language), which allows a grammar to be defined for a document, with
4 meaning embedded in tags. The trend towards XML-based applications should facilitate a
5 shared ontology structure, and allow metainformation to be associated with information
6 and describe data. XML allows intelligent integration of data from multiple databases.

7 Alternatively, we can allow data to be stored in heterogeneous formats across a
8 single unified database, within an object-oriented infrastructure. Each data object has a
9 "wrapper" that controls access, and provides an interface for queries. When a query is
10 executed, the method is invoked, and the result computed with the method and the data.
11 This is perhaps more efficient than the aforementioned approach, in that efficient data
12 formats are query structures are retained.

13 3.1 Data Structure

14 In this section we describe the structure of the records in the SDI database. The
15 next section describes the interface that allows user agents to submit data and update
16 information.

17 The data repository has the following key features: *price rules* are associated with
18 information, to control access to information on the basis of information requested and
19 properties about the requesting agent; meta-information to specify additional information
20 about data records, for example has the information been certified by a third party.

21 We describe the abstract structure of data records in the data repository. The
22 semantics of the data field and the price rules depend on the type of information. All data
23 records, whatever the information that they represent, contain the following elements:

24 1. Owner ID.

25 The owner ID is a three-tuple (Public Key, SDI identity code, Remote address).

26 The public key is provided by an agent that submits data, and is one half of a
27 public/private key pair in a public key based cryptographic infrastructure. The public
28 key is used to provide authentication of the agent, in case it wants to amend the
29 record in the future. The agent can submit a message signed with its public key to
30 prove its identity. An agent can use a different public key for each alias that it
31 maintains within SDI. The SDI identity code is a unique code, generated by SDI for

each agent alias, and provided to an agent to allow the agent to access the eBank and other ancillary SDI services, for example to collect payments received for access to the data. The remote address is an (optional) contact address for the agent that submits the information, for example an email address to a pseudonymizing module that will forward email to the agent under a number of different aliases.

2. Record ID.

The record ID is generated by SDI, and is used in the case that a user creates a number of different data records with the same owner ID.

3. Data object.

As noted above, we allow data in SDI to be of different types, and a general data object might be represented in an internal coding that is not known to SDI. Such an object must provide an *accessor* function to allow queries to be performed. The query-execution module takes an object and invokes the accessor methods to perform a query. In the most general form, a data object is represented as a three-tuple: (data type, data field, data accessor). The data type specifies what the type of data is, for example is it a user profile or a business rule. The data field specifies the data, and can be a private record that is only accessed by the accessor, which provides an interface to allow queries of the correct format for the data type to be performed.

In a simple special case, with data types designated by SDI centrally and data indexed with fixed fields, a data object is more appropriately represented as an association list between field names and values, e.g. (Field1, Value1), (Field2, Value2), etc..

4. Price rules.

The price rules provide the owner of information with control over the type of queries that can be performed on a particular data record. The price rules compute an ask price for a proposed query, on the basis of three pieces of information: what type of information, and what accuracy of information is requested; what certificates can the querying agent present; can the information be provided anonymously or must the identity of the owner of the information be revealed? As special cases, a price \$0 corresponds to "access is possible for free", and a price \$infinity corresponds to "no access is possible". The unit of currency need not be US dollars, but can be any unit

of currency, or as we also suggest in one useful variation “Community Dollars” that are SDI-specific dollars which can only be spent with vendors that are registered with SDI, and can also be restricted in various ways. As a special case, the prices allow a simple capability-based security system, because they can be set to \$0 or \$infinity on the basis of certificates owned by an agent that requests information. The price-rules are described in more detail in the next section.

5. Meta-information

The meta-information associated with a data record is expandable, but at present we suggest the following pieces of information: is the data is randomized, and is the information certified? The method of random data perturbation is fully described later in this document. Certification can be provided by third-parties, who can verify for example the age or nationality of an individual represented with an electronic profile.

We provide more information on metainformation below.

3.2 Price Rules

The price rules that can be associated with a data element are described in the next main section, the *Query Execution Module* section. The user associates a rule with data that the system uses to compute the price of a query during query execution. The price is further used within SDI to operate a *data marketplace*. Again, this is described in the next section. The user will eventually pay the *total price for its access to all data elements used to compute the final response to a query*; for example negative responses do not incur a price, it is only data that actively makes it into the process of computing the result of a query that matters.

We describe in the *Query-execution* module section the methods that are used to determine which data elements are used to compute the final response made to an agent, and therefore to compute the sum price for the final query. An agent is not charged for every access to data elements made during intermediate stages of executing a query. For example, if a query requests information about ten data records with high value associated with a particular field, the price of the query is the cumulative price for the access to the ten data records returned in response to the query, and not for all the data records queried in determining the ten records to return.

1 of field A, B and C', split into queries 'what is the value of field A', then 'what is the
2 value of field B', then 'what is the value of field C'? The pricing rule might state:

3 \$0.1 for any one of A, B or C

4 \$0.2 for any pair of A, B and C

5 \$10 for all of A, B and C.

6 In this case, by tracking the data record pointer with which a data record is accessed, the
7 price of the first two requests can incur an incremental charge of \$0.1, while the price of
8 a third request can incur an incremental charge of \$9.80.

9 At a per data-record level the price of a query depends on the response made to
10 the query. This is important because an affirmative answer to the question 'are you the
11 president of the united states' carries more information than a negative answer. This has a
12 slightly undesirable side-effect, in that when estimating the price of a query in the pricing
13 module, the *estimated price of a query might leak information about the result, even*
14 *without performing the query*. However, we believe that the benefits of linking price to
15 information content in a query outweigh this potential loss in value of information. Note
16 in particular, that in general the ask prices associated with data records are private
17 information and independently set, and therefore are not very revealing, especially within
18 a competitive market place.

19 3.3 Data Submission/Update Methods

20 In this section we describe the basic methods to register with SDI, submit data, update
21 data records, and access payments collected by SDI as information is queried. We use
22 "agent" to refer to the computer system that interacts with the central SDI data
23 warehouse, submitting data and requesting payment from the eBank. Agents may
24 represent individuals, vendors, or other self-interested parties.

25 Data records can be submitted under multiple aliases by a single agent, to provide
26 an additional level of control and flexibility in managing data submitted to SDI. For
27 example, if an agent represents an individual that is an online consumer, interacting with
28 different types of vendors, then perhaps the agent will use two aliases: one for while the
29 consumer is at work, and one for while the consumer is at home. Alternatively, the agent
30 can maintain a number of identities for different activities, or interests of the user. In the
31 preferred implementation agents, for example web-browser based client agents, can

1 manage an agent's selection of identities as it submits data to the central SDI warehouse
2 local to the user. The ability to submit information under multiple identities protects the
3 ability of an individual to prevent another agent building a complete picture about its
4 preferences and profile, while still allowing that individual to leverage as much of the
5 value associated with its information as possible. The SDI proxy agent, situated on user's
6 client machines, implements this functionality. It is described later in this patent.

7 A standard cryptographic public key /private key infrastructure provides a useful
8 technique to implement a system in which agents can maintain different aliases. As
9 proposed in the work of D. Chaum [Chaum 81; Chaum 85; Chaum91] a public key (PK)
10 and private key (SK) serves a number of purposes. First, the public key acts as an
11 identifier for the alias, a name. Second, the agent can compute a new private key/public
12 key pair when it requires a new alias. The agent keeps the private key secure, and this
13 provides a method to allow the agent to validate its identity, for example by
14 cryptographically signing a message with the private key. The signature can be verified
15 with the public key, and the keys can be selected with enough bits to make falsification a
16 computational impossibility. This infrastructure is outside of the current patent, but
17 standard in the art.

18 As described in the JANUS/LPWA system [BGGMM 97; BGGMM 98; GGMM
19 98], it is also possible to associate a public/private key pair with a pseudonymous e-mail
20 address, to allow information to be pushed to an agent that owns information under its
21 alias. The system is implemented via pseudonymous proxies which a user agent to poll
22 and check for new messages.

23 An alias can also be completely *anonymous*, but in this case the value of the data
24 provided may be less in the internal market place implemented within SDI (in the query
25 execution module). The convention for an anonymous alias is that the user agent
26 continues to provide a public key, and use a private key for validation of its identity.
27 However, in this case the public/private key is one-off and just for this data record, and
28 no return address is provided. An agent that submits information anonymously can still
29 recover payments from the eBank.

30 When a data record is first created SDI returns a data ID, so that the agent that
31 submits information can specify a particular data records in future, in case it creates a

number of data entries in the central SDI data warehouse. Submitting a new data record to the SDI data warehouse is accomplished with the following semantics:

(SDI identity code, SDI record ID) = SUBMIT(alias, data object, price rule, metainformation).

The following protocol is followed in the SDI data warehouse in response to a SUBMIT message:

1. SDI first checks the alias against its record of existing aliases. If the alias exists, then SDI first verifies that the alias is not anonymous (in which case it should only have one record), and then challenges the agent to sign a random message with its private key—to validate its identity. Once validated, the SDI identity code is returned to the agent.

Otherwise, if the alias is not found, then a new SDI identity code is created.

2. SDI then creates a new data record, with the data object, price rule, and metainformation provided by the agent, and then computes a new record ID, which is also returned to the agent as proof that the record has been created. This record ID is used to change the data in the future.

The owner of a data record can change the record with the following rules:

Ok = CHANGE(alias, record ID, data change)

Ok = CHANGE(alias, record ID, price change)

Ok = CHANGE(alias, record ID, metainformation change)

As with the SUBMIT command, first the SDI data warehouse checks that the alias exists, and challenges the agent that submits the CHANGE request to sign a random message with its private key, to validate its identity. Then the record ID is located, and again it is verified that the record is owned by the agent with the alias. Finally, changes are made to either the data record, the price rules, or the metainformation. If everything checks out, then SDI returns TRUE, otherwise SDI returns FALSE.

Finally, an owner of data can remove a data record with the following command:

Ok = REMOVE(alias, record ID)

The checks on the identity of the agent that requests that a data record be deleted are made as for the CHANGE command.

3.4 Meta-information

An agent can associate meta-information with a data record, that can serve a number of different purposes. Data records can be tagged with meta-information, that can include but is not limited to, Has this information been randomized? Certificates relating to the data record. For example, the meta-information might be a certificate from a third party about the integrity of information, or a certificate from an agent's client-side SDI proxy agent that the information in the record is unique, and not submitted under any other aliases by that agent.

A key example of the role of a certificate is described later in the patent, where we explain how the client-side SDI proxy can provide a certificate to state that this is the only data record with information X. This is useful, because it allows an agent that submits information to maintain multiple records, but still provide a guarantee to querying agents that certain valuable information is not duplicated across multiple records.

The meta-information may be associated with particular fields in the data associated with a data element, for example specifying that a particular piece of information has been randomized, or that a particular piece of information is highly sensitive and should be randomized before release.

4. Query Execution Module

The query execution module is a key component of the SDI system. Its key functionality is:

- a. Implement an internal market for information as queries are executed, ensuring that information that is provided for a lower ask price is used in preference to information at a higher ask price.
- b. Compute the price of queries based on information finally provided in response to a query, by keeping track of the data records that are used to compute a result.
- c. Interface with the pricing module to allow the price of a query to be computed without executing the complete query.
- d. Interface with the data repository, performing queries either on fields with fixed index labels (i.e. within a traditional database framework), or with respect to an

4.1 Price Rule Semantics

Each data record has an associated price rule that controls the price of accessing information. A data record can contain a number of different pieces of information, and data can be provided to different degrees of accuracy, so the price rule can be quite complex in general. In particular, consider a data record that represents the profile of an individual. The individual might be happy to have information released about some fields, for example its ZIP code, or its recent book purchases, but less happy to have information released about different fields, for example its salary or social security number. Clearly, different pieces of information within a single data record require different prices.

Furthermore, a price rule for the data record as a whole might need to be superadditive across data elements, such that it becomes very expensive to request too much information about data associated with the same user. One of the driving concerns behind the present invention is that at present it is possible for on-line vendors to collect information about a single individual via “cookies” (identifying codes which are left on a user’s client machine), and form a portfolio of information about various activities and preferences of that individual.

While information about an individual might be acceptable in small amounts, in large amounts the same information can soon become unacceptable. Furthermore, even if users release information anonymously in response to a query, if a lot of information is released the identity of the user can be compromised. Every additional piece of independent information that I state about my profile identifies myself a little more clearly, and acts to distinguish me from the profiles of other individuals. We discuss this further in a later section on random data perturbation, which describes how random noise can be added to data to counteract this effect.

The basic idea is that the system of SDI allows a user to associate a price rule with every data record, that computes the price that a user must pay to execute a query over that data. We allow the price to vary, depending on properties about the requesting agent, the amount of information requested, and the level of identification that is required of the owner of the information in responding to the query.

The abstract form of a price rule for a data record is a function:

1 (Information-request x Certificates x Level-of-identification) \rightarrow Price

2 Given a request for information, a set of certificates that a requesting agent can present,
3 and the level-of-identification that is required of the agent that provides information (i.e.
4 anonymous, or revealed identity), then a price rule computes an ask price. This is the
5 price that an agent must pay to execute the query on the data record.

6 However, as we describe in the next section, an agent only actually pays this price
7 if the information provided is used to compute information in the final response provided
8 to a query. For example, consider query “Select all data records close to record X”. The
9 query is most simply executed by computing the distance between each record and record
10 X. Although a negative response from a record that is not close to X provides some
11 information about that record, the information is not used to compute the information (i.e.
12 set of records) that is finally provided in response to the query.

13 As in the representation of heterogeneous price rules in SDI, the invention of SDI
14 allows heterogeneous methods to compute prices for queries. In the most general case,
15 we allow a price rule to be computed as a price method, which is invoked for a data
16 record before a query is to be performed. The query-execution and pricing modules
17 simply invoke the price method, and the data object provided by an agent returns a price
18 for the query. In this most general version, the role of SDI is limited to providing the
19 price method with the query, the certificates of the requesting agent, and the level-of-
20 identification that is required.

21 In a simpler variation the system of SDI can provide a number of default price
22 rule languages, which allow a user to specify in simple but quite flexible terms a price
23 schedule for queries. We describe two such rule languages: an uncertainty-based additive
24 price rule; and a grouped additive price rule which is a simplified version of the general
25 uncertainty-based rule.

26 The uncertainty-based additive price rule allows a user to adjust the price for a
27 query based on information about the requesting agent, the amount of information
28 requested, and whether the information is required anonymously or with a revealed
29 identity. It does not allow a non-linear coupling across the price of information about
30 multiple attributes, but instead includes a simple upper-bound on the amount of
31 information that can be requested. Such a non-linear coupling could be added with simple

1 interaction terms between the prices on elements, for example if “information about
2 more than half of the elements in this set is released, then add price term L”.

3 **4.1.1 An Uncertainty-Based Additive Price Rule**

4 The uncertainty-based additive price rule allows a querying agent to present a
5 certificate to prove that it is entitled to query a user’s data for free, or prove that it is
6 entitled to receive data at a price. If a querying agent can present no certificate from
7 either set, then it is not allowed to query the data.

8 Then, a user can define two different price rules, one for data revealed
9 anonymously, and one for data revealed with an identity. A price rule computes a price-
10 term for each data element that is queried, with the total price for a query computed as the
11 additive sum of all single element terms. The price for a query on an individual element
12 is a linear function of the uncertainty with which the information is provided. For
13 example, if a query requests the exact value of an element there is zero uncertainty, if a
14 query requests a range of values for the element, or a ‘value + random perturbation’ there
15 is some residual uncertainty, and we allow a user to associate a cheaper price with this
16 result. The uncertainty is normalized with respect to the population of data elements in
17 the SDI data warehouse, e.g. an uncertainty of 10% indicates that the reported
18 information is consistent with 10% of data elements in the database.

19 The price-rule is parameterized with the following information for each data element
20 (or set of data elements, if a user chooses to associate data elements with a type):

- 21 • A set of certificates that allow a query to be performed for free.
- 22 • The price of an accurate response
- 23 • A function to compute a discount from the accurate price, based on the
24 uncertainty of the response.

25 There are many possible discount functions, that include but are not limited to the
26 following:

- 27 • A linear price discount function, such that 100% uncertainty corresponds to a
28 discount equal to the accurate price. In this case, the user must only provide the
29 price for an accurate response.

- A step price discount function, where the level of discount is associated with a set of discrete uncertainty ranges, e.g. <10% no discount, 10—30% receives a 20% discount, and 30-100% receives a 70% discount.
- A quadratic price discount function, where the level of discount is associated with a quadratic polynomial function of the uncertainty, in this case a user needs to specify a number of intermediate discount points and SDI can compute a best-fit quadratic function.

Stage 1: Present a Certificate (Price-discrimination)

First, the querying agent can present a certificate to allow it to query information for free, or to allow it to query the data for a payment. If the agent can present no such certificate, then it is not allowed to access the data.

Assume that the requesting agents falls into one of three classes: it presents certificates to make its access free, it presents certificates to make access possible at a price, it presents certificates to make access impossible (infinite price). Let Free denote the set of certificates that allow a query to be performed for free, and Charge denote the set of certificates that allow a query to be performed for some charge. The failure to present certificates in either class indicates that an agent is not permitted to query the information. We do not suppose that the presence of a certificate can indicate a negative property, because there is no method to force a querying agent to provide a certificate. We also assume that agents must present the same certificates to access all data elements. It is trivial to relax this assumption.

Stage 2: Specify the type of information required.

Second, the querying agent specifies whether it is required that the information be provided along with the identity of the owner of the information. Assume that the agent's identity is either revealed (with a real or pseudonymous identity), or anonymous. The price required for particular types of information will vary with the level of identification.

Stage 3: Select the appropriate price rule; Compute sub-prices.

The “cross-product” of the certificate (i.e. free, price, or none) and the level-of-identification (i.e. anonymous, revealed) triggers a price rule. We allow the owner of the information to associate a price rule for each combination of certificate and identification, and with each data element. Many simplifications, for example “no access with no

1 certificate” and labeling data elements in a class, and providing a price rule for a class of
2 data elements are possible.

3 Without loss of generality, we can describe the price rule for a particular data
4 element, and a particular cross-product of certification and identification. The price-rule
5 contains a base-price, which is the price for accurate information about the data element,
6 and a discount-factor, which reduces that base price by an amount related to the level of
7 uncertainty in the response.

8 A query might permit a response that includes a level of uncertainty, for example:

- 9 • Is the value of the data field greater than a threshold value?
- 10 • What is the range of values in which the value lies?
- 11 • Compute an additive noise perturbation, and respond with a “value plus noise
12 perturbation”.

13 The randomized perturbation technique, which allows the agent to return randomized
14 information about its true value, is disclosed later in this patent. For simplicity, we can
15 assume that any non-exact query can be associated with an uncertainty-value, which is
16 computed based on the proportion of the population of data elements that can have values
17 within the accuracy of the response. The system of SDI can compute this uncertainty-
18 value, perhaps approximately based on static statistics computed on a periodic basis over
19 information in the database.

20 For example, SDI might assume that values are uniformly distributed between the
21 lower and upper values over the population of data elements, and compute uncertainty as
22 the fraction of values within the range indicated in an agent’s response. Suppose that all
23 data records have values between 0 and 100, then a query ‘is x > 70’ has residual
24 uncertainty 30/100 (normalized) if it receives a positive response or 70 /100 if it receives
25 a negative response, while a query ‘what is your value of x’ has zero uncertainty for all
26 responses.

27 Assume that the uncertainty value is between 0 and 100%. The discount factor,
28 discount(X), is computed as a function of the uncertainty X, with a larger uncertainty
29 leading to a larger discount.

30
$$\text{Price}(Q) = \text{Base-Price}(Q) - \text{discount}(\text{Uncertainty}(Q))$$

- Low price, if an agent presents a Charge certificate, and requests information with a High degree of uncertainty.
- High price, if an agent presents a Charge certificate, and requests information with a Low degree of uncertainty.
- Infinite” price, if an agent cannot present a Charge certificate.

Different price functions can be defined for anonymous and revealed-identity. The price rule has the following components:

- a) A set of certificates that allow data access for free.
- b) A set of certificates that allow data access for some charge.
- c) A list of price-functions, each defined with two prices for revealed-identity and two prices for anonymous identity. The pairs of prices are for accurate information, and approximate information (with a minimal level of approximation).
- d) A mapping of data elements to price-functions.

Again, the price of a full query is computed as the sum price over all information requests to all data elements.

4.1.3 Superadditive Price Rules

We can allow a non-linear interaction between the price of queries on individual data elements with the introduction of additive interaction prices, which are fired when a threshold is exceeded for the total amount of information released over all data elements accessed in the same query. This is a simple approximation to a more general combinatorial price rule, which would price all combinations of data access terms explicitly. For example, a rule could state: “If more than T data elements are accessed in set Sensitive then add a “bundle price” L to the total sum price of the query.”; or: “If the total weighted access to data elements in set Sensitive exceeds threshold T_1, then add bundle price L_1 to the total price of the query”; where the weighted-access is computed with weights equal to the degree of uncertainty associated with queries on elements.

4.2 Computing the Price of a Query

A query can be quite complex, involving a number of operations on data stored in the data repository. Call a query session a sequence of operations performed in response to a query, and before the answer is returned to the agent that submits the query.

1 separation is maintained between the data and the state of the program that executes
2 queries.

3 Each query session is associated with a table of data record pointers. The table
4 records valid data record pointers, that can be referenced in later query operations, and
5 the current price of queries performed with the data record:

6	Data Record Pointer		Current price
7	-----		-----
8	P_1		\$0.10
9	P_2		\$0.20
10

11 The SDI query execution module performs another key book-keeping service, recording a
12 list of data record pointers that have been used to compute a result in the program that
13 executes a query.

14 For example, suppose that set PS contains a list of data record pointers, and a
15 subsequent query operation makes the following command:

16 `_count = Count(PS, 'property X')`

17 which can be interpreted as “count the number of records in set PS with property X”. The
18 query-execution module maintains a table of variable accounts, which records the data
19 records that were used to compute a result:

20	Variable Name		Associated data record pointers
21	-----		-----
22	_count		P_1, P_4, P_6
23	_tmp		P_3, P_8, P_10
24

25 Whenever the value of a variable is computed from the values of a number of existing
26 variables, then the list of data record pointers is the union over the data record pointers
27 which were used to compute the values of the other variables.

28 Eventually, when a result is reported at the end of the execution of a query, the
29 price of the query is the sum of the requested payment for the data record pointers that
30 were used to compute the information provided, as stored in the table of data record
31 pointers.

1 Figure 3 illustrates the method diagrammatically. We maintain a “firewall”
2 between data and the state of the query execution, with all access to the values in data
3 records via data record pointers, and invoking price rules associated with data records.
4 On the query-program side we maintain a table of data record pointers, and the cost of the
5 cumulative information received from that data record under that pointer, and a table of
6 variables and the data record pointers which were used to compute the values in the
7 variables.

8 All manipulation with the values of variables on the “query-side” of the data
9 firewall can be performed with no additional data-access charge, because data access
10 charges are only incurred in computing values of variables, when that directly or
11 indirectly requires that additional information be collected about the data records.

12 The only charges are incurred when the processor receives information. This
13 information might be in the form of a number, that represents the count of records with a
14 particular property, aggregate information in the form of a new collaborative filter, or raw
15 data extracted from data records.

16 For example, SELECT operations are free until information in the data record is
17 used to compute the value of something which is used in computing the final response
18 returned in answering a query. A charge is accounted for against the data records which
19 are queried and selected, but that charge is only levied if information is used.

20 **4.1.1 Internal Market within the Query Execution Module**

21 The data side of the firewall is also responsible for performing the cheapest queries
22 possible. For example, if there is a request to retrieve 100 data records with property X,
23 then the database retrieves the data as cheaply as possible based on the access-control
24 rules of the owners of the information in the database.

25 All operations are implemented within an internal market. The internal market ensures
26 that information is provided according to the price rules associated with data records, and
27 also places providers of data in competition with each other.

28 The system of SDI implements an internal market, with queries executed as
29 cheaply as possible. Both the SELECT and the PROCESS commands allow an associated
30 restriction on size, e.g. “SELECT 100 records with property X”, or “PROCESS 10 of the
31 records with data pointers in list X1 with method Y”.

1 In one greedy variation, each stage of a sequential query is executed as cheaply as
2 possible. For example, for each operation the query is implemented by choosing the
3 cheapest data records that satisfy the criteria. For example, the records that are selected
4 are the 100 cheapest records to query that satisfy property X, and the 10 records in X1
5 that are cheapest to process with method Y are selected.

6 This greedy method is not necessarily optimal for a sequence of queries, for
7 example when the records that are cheapest for the current query are more expensive for
8 a later query. Another variation allows a set of records to be maintained, with the
9 cheapest records used in reporting the final result to the user. It might not be optimal to
10 use the cheapest records at every point. For example, if a query is described as "SELECT
11 100 records with property X", then "PROCESS with method Y", and "RETURN result",
12 it can be cheaper to select more records (e.g. 1000) initially, process all the records, and
13 then return the 100 records that are cheapest after the processing.

14 The method of storing the data record pointers in a table, along with the price
15 required for information, and keeping a record of the data record pointers that are used to
16 compute values during the execution of a query program provides a technique to compute
17 the total price of a query.

18 In the previous section we described price rules, which are associated with data
19 records and prescribe on a record-by-record basis the price of executing a query.
20 However, only data records that are used in computing the final result that is returned to
21 an agent at the end of a query receive payment.

22 **4.1.2 Query Optimization**

23 Finally, consider the following two examples:

24 COUNT the data records with PROPERTY X
25 and COUNT the data records with PROPERTY (not X)

26 In an optimal pricing model these queries both have the same cost. To focus ideas further,
27 suppose that no data records have property X, such that without additional query
28 optimization the cost of the first query is \$0, while the cost of the second query is the sum
29 cost over all data records of reporting that they do have data property X.

30 To solve this problem we add a simple query optimization routine in the pricing
31 module, which can test different formulations of properties (i.e. negatives), and the effect

1 of reformulation on price. The order of operations for a given expression of properties
2 does not change the price, but the way in which properties are expressed can affect the
3 price.

4 In another extension, we can use statistical methods to optimize performance. For
5 example, a query 'what % of data records have property X' can be computed using a
6 sample of the total data records, for an accuracy/price tradeoff. This functionality can be
7 offered within the pricing module.

8 Another key technique for query optimization is to allow the SDI central server to
9 perform pre-computation, for example to classify data records into useful bundles, and
10 allow future queries from agents to limit themselves to bundles of data records that would
11 appear to have useful properties. The cost is then incurred for further information
12 revelation from these records, and avoided for information revelation from records which
13 may not have useful properties.

14 4.2 Example of Query Execution

15 As an example, consider the following composite query:

- 16 1- SELECT all data records that relate to users that have purchased a flight in the past
17 month,
- 18 2- PROCESS: Sort the data records according to salary,
- 19 3- SUBSELECT: Limit the records to those users that are willing to have their identities
20 revealed,
- 21 4- RETURN RESULT: Return the identities of the remaining users with the top 100
22 salaries.

23 In step 1 data records from the SDI data repository are selected which have the
24 required property, in this case corresponding to profiles of individuals who have
25 purchased a flight in the past month. The table of data record pointers is augmented with
26 temporary pointers (valid for this query session) to records with this property, and the
27 price charged by each data record to reveal this information. Notice that the table only
28 includes pointers to data records that have this property. Although data records that
29 respond negatively to the query reveal information to the query-execution module, that
30 information is never revealed to the agent requesting the query, and therefore no payment
31 is received.

1 In step 2 the data records that relate to profiles of users that have recently
2 purchased a flight are now sorted in terms of their salary field. An algorithm to sort
3 performs a sequence of comparisons between data records, with each comparison
4 requiring that more information is revealed about the underlying salary of the profile, and
5 changing the price associated with the data record. The prices are updated in the data
6 pointer table according to the data record price rules, and the query program now has a
7 list of ordered data record pointers.

8 Now, in step 3, the query program initializes a new set of data record pointers.
9 This includes all the pointers for users that are willing to have their identities revealed.
10 The pointers are maintained in order of salary. At this stage the entry for this new
11 variable in the table of variable accounts only associates the data records with this new
12 property.

13 Finally, in step 4, a new variable is initialized to contain the first 100 data record
14 pointers from step 3. These are the data record pointers listed in the variable accounts
15 table. These are the only data record pointers that receive payment as a result of this
16 query, because the intermediate information provided by other data records remains
17 within the SDI query execution module and is not released to the agent that performs the
18 query.

19 4.3 Discussion

20 The pricing model as outlined above has the following attractive property:

21 The price of a query is independent of the order in which operations are
22 performed, and optimization of order is not necessary for reasons of price.

23 This is nice, because it means that agents do not have to perform lengthy optimization to
24 structure queries in cost-efficient ways. The pricing module also ensures that:

25 Agents only pay for information received as the result of executing a query.

26 Finally, we implement an internal market, so that with all other things equal, data mining
27 and other types of queries are executed as cheaply as possible:

28 An internal market favors data records with low cost pricing instead of records
29 with high cost pricing, with all other things equal.

30 As an example, consider this query:

31 COUNT (SELECT 'FIELD1 = X' and 'FIELD 2 = Y')

1 This query can be structured as 'Count (Select(X and Y))', or 'Count(Select(X,
2 Select(Y)))', or 'Count(Select(Y, Select(X)))', all with the same cost. Although different
3 amounts of information is requested from all data records in computing the query in the
4 three different formulations, the final cost is computed only over the records that have
5 property X and Y, and these records are the same across formulations and have released
6 the same information.

7 Similarly, consider the query:

8 SELECT(Z, Select(Y, Select(X))); return '2'.

9 i.e., perform some complex nested query but return the fixed value '2'. The query
10 execution module tracks that no data pointers are used to compute the value returned to
11 the agent, and the querying agent is not charged for the information provided about data
12 records to compute the SELECT operations, because that information is retained within
13 the central SDI data server.

14 In a SELECT query the agent can only ever pay for information provided about
15 data records that respond positively to a SELECT, because no information is ever
16 released to agents about the data records that respond negatively.

17 In another example, consider a query 'COUNT the data records with PROPERTY
18 X', that returns the result 'none', i.e. there are no data records with this property. This
19 query is priced at \$0 because none of the agents that own information provide
20 information to compute the

21 Therefore, a query of the type 'how many data records have type X' that answers
22 'none' has price \$0, because the individual agents have not revealed any information.
23 Similarly, because we would like all equivalent queries to have the same price, a query
24 "how many data records do not have type X" in the same database (answer "all of them")
25 should cost \$0 because the same query could have been formulated the other way. A
26 reasonable way to ensure that a query is formulated in the cheapest possible way is to use
27 an query optimization manipulation program, of the type that is common in database
28 research, during the 'price quote' phase in order to find the formulation with the cheapest
29 price.

5. Pricing Module.

The pricing module is a core component of the central SDI data warehouse. It has two functions: (a) to *predict the price of a query without executing a complete query*; (b) to aid in the allocation of payment given a price quote.

We build into the pricing module a method to allow an agent to make a quality/cost tradeoff before executing a query, and even to specify a restricted query that will respect an agent's budget. Already embedded in the query-execution module is an internal market for data records, to ensure that cheaper data is used in preference to more expensive data access, given two identical data records. Furthermore, already embedded in the query-execution module is that agents only pay for the *provision of information*. Finally, the method computes a price for a query which is independent of the order of operations.

We propose the following method for providing initial price quotes to agents in SDI:

1. Compute a range of price quotes in response to a query, for different amounts of computation (which can be varied by limiting the input set of data records), and for corresponding values of a "quality metric" provided by an agent that submits a query.
2. Allow the querying agent to select an optimal price/quality/computation tradeoff, and then receive payment, execute the query, and provide the response.
3. Allocate payments to accounts in the eBank that correspond with data records which are used to compute the value of the result of the query, with payments adjusted if necessary to account for any error in the price quote.

5.1 Computing a Price for a Query

We allow a small *fixed price* F for an initial price quote, to cover the cost of computation and prevent agents saturating the server. The price module computes a *price quote* based on a statistical technique to compute the price of the full query without executing the full query.

The basic idea is to perform the query on a number of records, e.g. 1%, 5%, and 10% of the total records that will be used for the full query, and then extrapolate to the full query size. We need to perform the complete query because on a subset of data records, instead of part of the query on all data records, because query programs can have different phases—and each phase might incur very different information costs. The

number of records that are used for the query can be restricted by making a random sample of the total domain of data records used for a full query.

5.2 Allowing a Price/Accuracy Tradeoff

In computing the initial price quote we also suggest a tradeoff for the user between “result quality” and cost. *Quality is subjective, and usually best measured by the agent that requests a query on the data. Therefore, we suggest that the querying agent should provide the metric, but limit this metric to a scalar value to prevent information leakage.* We allow an agent in formulating a query to specify a key metric that it will use to choose an appropriate size of query to execute.

The core technique that we use to control the accuracy of the result of a computation is to limit the number of total data records that a query runs over as input to the query. Take a random selection of all the data, and use that for the processing. At this initial step we limit the price and the amount of computation performed.

As an example, consider a query to collect pseudonyms for agents with useful properties. The metric of interest in this case is the average per-pseudonym cost, and this can be provided by the pricing module *without providing any useful information to the querying agent.* Given this information the agent can then decide whether to proceed with a query, and how much money to allocate.

Similarly, perhaps a querying agent cares about the quality of match between data records and data records with ideal properties. The agent can provide a method to *instrument* its query to compute this quality, and the price module can in pricing the query provide guidance about the tradeoff between running the query over different numbers of data records.

Another more involved method is to introduce randomization as the query is processed, because some owners of information may provide randomized information more cheaply than exact information.

That will vary depending on the number of data elements that are queried.

5.3 Making a Price Quote for a Budget

Consider the problem of an agent that wished to perform a query with a budget ***B***. For example, I would like to spend \$1000 to compute a collaborative filtering model. There is a particularly simple way to allow this:

1 First, estimate the cost of building a collaborative filtering model on the basis of
2 N data records, selected at random (or with some preprocessing of the database to
3 identify useful data records). Estimate price P for the query.

4
5 Second, adjust the number of data records used to compute the actual query based
6 on the price P that was computed, and the budget B . The number of data records,
7 assuming a linear model of per-data record charging is simply $N \times B/P$.

8 If there is reason to believe that the cost of a query is a non-linear function of the
9 number of data records, then a number of samples could be priced, for suitable sample
10 sizes N_1, N_2, \dots etc.

11 Ultimately, the result is a query that is adjusted in scope by an upfront limit on the
12 number of records used for the query domain. The price module can scale the cost of the
13 query, and make an appropriate price quote.

14 15 5.4 Optimizing the Price of a Query

16
17 We also suggest a technique to adjust the formulation of properties in a query, e.g.
18 'COUNT(SELECT records with property X)', could be reformulated as 'Count(all
19 records) – COUNT(SELECT records without property X)', possibly for a smaller price.
20 A random search technique can provide one simple approach to optimize query formulas.

21 22 6. Communication Module

23
24 The communication module in the central data warehouse takes incoming messages from
25 agents and decides how to handle them, and sends responses to queries back to agents.

26
27 The action selected in the communication module depends on the *type* of the message, for
28 example the message can be a data-update request, a new query to execute, a request to
29 add new profile information, a request to find agent profiles that match, or a request for
30 processing a set of profiles that meet a particular criteria with an application in the
31 database.

Figure 4 illustrates the communication module in block diagram form. Messages are received from agents, and passed to the data repository in the case of data-update commands, or requests to add new information to the data repository. Alternatively, if the message represents a query then it is either placed in the queue for execution, i.e. the pending queries queue, or in the set of persistent queries—which are periodically checked by the query execution module and executed when the correct conditions exist in the data warehouse. The communication module can be implemented using standard message-handling methods known to those skilled in the art.

7. Physical Location of Data

The default implementation of SDI assumes that information is physically stored in the central SDI data warehouse. However, this is not necessary. The system of SDI is designed to present a privacy-protected marketplace for data, where queries can be executed with complete transparency as to the physical location of the data over a network. It is not necessary that data is physically uploaded to the central SDI data warehouse. Instead, we allow a user of SDI to provide a *virtual pointer* to the location of the data on another machine, for example hosted by a trusted third-party, or for example their own data server when that server is located within their own facility and always online. Alternative data locations provide tradeoffs between computation efficiency, communication cost, delays, and privacy.

7.1 The Convergence of Network Devices

The invention of SDI is in no way limited to a centralized implementation. The invention is characterized by its methods and protocols for handling personal information and query execution. As best as possible the invention applies to any distributed system implementation that follows the disclosed methods.

1 consists of a remote shared terminal (e.g. ATM kiosk, POS device), personal data may be
2 transmitted from the user's own client device if/when and at the time it is needed and
3 authorized by the user. This approach would be certainly feasible within the "always on"
4 client scenario, for example with a user's data stored at his/her set-top box, which is
5 always "on-line".

6
7 The devices with which a user interacts can check a user's data use policy, and then push
8 data accordingly to the SDI network servers. Because servers at the network level are
9 inherently not as secure as the user's own client, extremely sensitive personal information
10 may still be stored even at a light client, i.e. data can be split across multiple locations. In
11 general the user may make a tradeoff between data management (latencies, bandwidth)
12 costs and with the potential loss of information that can occur when data is released to
13 other parties, even if that party is the central SDI data server.

14 15 **Infrastructure for Smart Home / Smart Office Intelligence**

16
17 The management of the potentially vast amount of information generated by such
18 devices, as well as its privacy protection, is enabled by the "personal agent", an on-line
19 software entity that represents and acts on the behalf of the individual user. Furthermore,
20 this agent is capable of communicating with and negotiating terms with agents
21 representing other entities, including systems that form the infrastructure backbone for
22 intelligent homes and offices.

23
24 Such systems (as embodied by the LEIA disclosure) have access to a broad range
25 of input sources, allowing them to monitor different aspects of the user's activities or
26 location. Inputs can include data derived from:

- 27
28 * Email, wireless PDA, telephone communications.
29 * Web browsing
30 * Personal Computers
31 * Personal Digital Assistants

- 1 * Pagers
- 2 * Wrist watches (typically one-way, receive only)
- 3 * ATM machines, point of sale kiosks
- 4 * Ceiling-mounted video cameras
- 5 * room-based motion and heat detectors
- 6 * Internet-enabled automobile
- 7 * Chip-enabled appliances
- 8 * Electronic Schedule Books

10 **Smart Home Application**

11

12 The concept of a “smart home” is not a new one: Even in the 1950’s futurists

13 discussed the possibility of imbedding intelligent systems into a house, relieving the

14 homeowners of such boring daily tasks as watering plants or adjusting the thermostat.

15 Such a system could easily be implemented today using control computers running

16 simple sets of IF-THEN rules, but might soon prove unreliable because of its inability to

17 adapt to new situations. For example, a garden party could easily be ruined if

18 inadvertently scheduled at the same time as a scheduled lawn-watering.

19

20 LEIA provides a much more sophisticated alternative. Given a house that

21 incorporates an internal local area network (LAN) linking household appliances, sensors,

22 communication devices, and home computers, LEIA could easily form the basis for a

23 truly intelligent and adaptive home environment capable of handling daily chores,

24 security, and information delivery.

25

26 Firstly, LEIA could easily subsume the tasks traditionally slated for use in a smart

27 home; however, its ability to infer the location and needs of family members would

28 greatly enhance the handling of these tasks. For example, the above-mentioned garden

29 party would not be spoiled because a motion sensor would inform the watering system

30 that a party is taking place on the lawn. Or, being set in a power-saving mode, LEIA

31 could light only those rooms currently being used; as a family member moved around the

1 house, lights would be turned off and on, tracking the person's movement so that only
2 their immediate surroundings are lit.

3
4 Secondly, LEIA could use its ability to predict informational needs (in a manner
5 very similar to that used for the Automotive Information Delivery System) in the house,
6 customized to the personal tastes of different family members. One could imagine recipes
7 being delivered audibly in the kitchen, or a television program that follows a family
8 member from room-to-room as he strolls around the house.

9
10 Finally, LEIA would have the sophistication needed to deal with intelligent
11 appliances, telling the swimming pool to clean itself, for example, an hour before the
12 commuter returns home.

13
14 LEIA provides a unified infrastructure for the increasingly sophisticated systems
15 being incorporated into modern homes. It would act as the central manager, making sure
16 that basic household tasks, security, and informational needs all work in a coordinated
17 and intelligent manner, adapting smoothly to the uncertainties of daily life.

18
19 A smart home would have several different types of sensors:

20
21 i) "Passive" detectors --

22 *motion detectors

23 *sound detectors (including voiceprint and speech analysis module)

24 *Active Badge detectors

25 *Window and door sensors (are they open/closed?)

26
27 ii) sampled communications

28 *telephone

29 *e-mail

30 *radio/TV

31 *computer usage

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

- iii) appliances
 - *alarm clock
 - *light switches
 - *stove
 - *dish washer/clothes washer
 - *water taps

In addition, accessed via secure channels (perhaps using a personal agent as an intermediary), three major types of personal data are also useful for running the smart home:

- i) Explicit rules for routine tasks (E.g., desired schedule for lawn watering)
 - ii) Expressed preferences (E.g., a desire for morning news about international events)
 - iii) Historical database of previous interactions between the family and the house
- (This will allow LEIA to learn daily routines and family habits).

There are two major classes of data that need to be served by LEIA: (1) control signals for household systems and appliances, and (2) information desired by household members.

To a large extent, household systems can be appropriately controlled using a traditional rule-base, although a fuzzy rule-base would be inherently more flexible (IF <the bedroom is somewhat chilly> AND <it's almost time to wake up> THEN <substantially increase the bedroom's heat>). Many of these rules would be explicitly programmed by the family from the beginning; others could be learned using inductive methods.

1 Although most of the household systems could run unsupervised using such a rule
2 base, LEIA would remain the ultimate arbiter, and could override the rule-based system
3 in exceptional cases. For example, if the family suddenly has to go out-of-town, LEIA
4 can infer their absence and inform the household systems to go on energy-saving modes.

5
6 It should be noted that certain state variables will have a major impact on this
7 analysis. The date, time, mood (inferred from sounds levels and voice content), and
8 current composition and location of people in the house should be carefully conditioned
9 on. Strong daily routines will emerge (for example, a typical working day), but
10 exceptions will also be noted (when the teenaged son is home alone, and in an exuberant
11 mood, play pop music LOUDLY).

12 13 **Extension to Data Delivery**

14
15 Given the ensemble of sensors that form part of the smart home, it would not be
16 difficult to monitor the types of off-line information being consumed by the inhabitants.
17 Biometrics (systems monitoring body state, facial expressions, retinal reactions), sound
18 sensors (capable of translating speech to text, and then text to content, using natural
19 language processing techniques), and video monitors (which could observe reading
20 material such as magazines, textbooks, tv viewing guides, cookbooks, etc., and again use
21 NLP methods to scan for content), would all form part of the input set.

22
23 This information is combined with all other sensor readings to infer the general
24 focus of user interest. Within the scope of the present network architecture, the parent
25 patent application describes how such contextual data can be converted into target object
26 profiles and through the use of the profile processing module “similar” target objects may
27 be retrieved from the electronic medial distribution system (cached locally or over the
28 network) and delivered and displayed at the local user terminal device for viewing by the
29 user. Such media (which may be tailored for delivery at a RV set-top) may include
30 streaming video/audio, multi-medial files or video-on-demand which are delivered in
31 near real time upon activation of the request and which are of matching similarity (or

1 complementarity) to the print media viewed by the user presently. Alternatively, the data
2 captured by the miniature camera device relating to the particular pages of print media
3 presently observed by the user may consist exclusively of an identifier code a portion of
4 which is used by the system for purposes of identification of the code (or alternatively
5 such identification may be determined by its physical coordinates on the page itself). The
6 other portion of the code is a unique identifier of the particular page, which may be used
7 to identify manually by a content expert whereby select print materials and particularly
8 their associated electronic counterparts are either matched, designed and/or customized
9 for use within their system context of the present convergent media application. It is
10 apparent that such a system could also be usefully deployed to benefit advertisers, e.g.,
11 presenting advertisements electronically to correspond with particular printed content or
12 matching printed ads with corresponding electronic counterparts.

13

14 This complimentary content, be it audio, video, text, or combinations thereof,
15 including advertising conforming to user-defined conditions, is delivered through various
16 in-home devices located in the proximity of the user.

17

18 **Smart Office Application**

19

20 Large business centers can be very confusing places: people are often away from
21 their desks, many different meetings are scheduled and rescheduled during a single day,
22 people are called off to emergencies, and clients drop in for unexpected visits. The
23 inability to coordinate people and information in such situations leads to greatly reduced
24 efficiency.

25

26 LEIA offers a unified solution in the face of this chaos; by linking together
27 people's schedule calendars, communications systems, contact information, and location
28 coordinates, a system can be devised to (1) shepherd a visitor (even if unexpected)
29 around the business center, (2) push vital contact information about the visitor to those in
30 the company who will be visited, and (3) adaptively coordinate important meetings
31 within the company.

1

2 When a visitor comes to the center, he will:

3 a) sign in at a front desk; rather than sign a book, he logs in at a terminal that signals his
4 presence to the company LAN.

5 b) receive a hand-held computer that, while presenting the visitor with useful
6 information, will emit a constant signal enabling LEIA to pinpoint his position at all
7 times.

8

9 On the employee side, LEIA can acquire signals from:

10 a) active badges

11 b) telephones

12 c) beepers

13 d) interactions with terminals

14

15 All sensors are connected directly to a LAN internal to the company. Only basic
16 security measures need to be enacted.

17

18 Registered visitors will likely be profiled in standard contact files containing
19 information about their own company, position, professional interests, previous
20 interactions with the company they're visiting, and basic personal details (e.g.,
21 photograph, wife's name).

22

23 Employees of the company using this LEIA-based system will have much more
24 information available. In addition to the standard contact information, LEIA will have
25 access to their phone and e-mail logs, daily calendar/schedule, work group and project
26 details, position within the company hierarchy,
27 and work-station notes (e.g., an unexpected "out to lunch" message could be posted on-
28 screen by the employee before they run off).

29

30 Although some higher-level intelligence is needed (e.g., for matching interest
31 profiles between employees and visitors), the office-based system could well be run with

1 a combination of a rule-base and optimization routine (which would be capable of
2 coordinating conflicting schedules). Since this version of LEIA will be installed in a
3 fairly controlled environment, the problems of location and timing will be fairly trivial,
4 the former solved by active badges, the latter solved by smart calendar agents.

5
6 *For the visitor:*

7
8 If the visit is expected (and previously scheduled), LEIA will double-check the
9 availability of the personnel to be visited and confirm or alter the schedule as needed. If
10 the visit is unexpected, or certain employees are unavailable, LEIA will take into account
11 the purpose of the visit and the visitor's professional interest profile to locate those
12 available employees who would be best suited for a meeting.

13
14 Once the visit schedule has been arranged (although it could be adaptively altered
15 during the course of the visit), the visitor will be fed information via the hand-held
16 computer they've been issued. This might give them a visual representation of the
17 building's layout, along with the best route to the next office to be visited (a more
18 advanced building might simply have computer displays in its corridors capable of
19 flashing directional signals to the visitor as he passes them). In addition, it might give the
20 visitor basic information about the person he's about to visit (for example, they might
21 have interests in common, or have gone to the same university).

22
23 *For the employee to be visited:*

24
25 Whether or not the visit was previously scheduled, LEIA communicates with the
26 employee's personal calendar agent and double-checks the employee's current location.
27 If the employee is at a fixed location and has some free time, the visit is scheduled.
28 Obviously, his connection to the visitor will determine what information, in addition to
29 the standard contact data, should be pushed to his workstation.

30
31 *For multiple employees requiring a meeting:*

1
2 Given that multiple employees have signed up for a meeting, or that a single
3 employee has sent out a signal indicating the need for a meeting and its priority (standard
4 or emergency), LEIA gets to work coordinating the details. Cell phone and beeper signals
5 give employee locations outside the office, active badges signal employees already at the
6 office. Calendar agents are contacted for availability; a standard meeting might depend
7 on people having time slots free, whereas an emergency meeting would be given priority
8 status, and could bump items already scheduled (given that they have been assigned low
9 priority).

10 An optimal meeting time and place (or teleconnection for remote employees) is set up,
11 and any pre-meeting information is forwarded to the appropriate workstations.

12
13 meeting, engaged in an important phone conversation

14 **Resolution Credentials**

15
16 Section 2.2.1 of this patent discloses the use of resolution credentials for the
17 facilitation of agent-agent interactions. Resolution credentials enrich the quality of such
18 interactions because they certify the status claimed by individuals, allowing agents to
19 maintain relationships of trust. On a more practical level, resolution credentials are of use
20 for matching and introducing users to each other on the basis of common interests, for
21 guarding access to high-level users (with the agent acting as a gate-keeper), and for flagging
22 times at which an individual is reachable (and not tied up in meetings, or the like, as
23 specified by the scheduling agent).

24 25 26 7.2 General Description

27
28 We allow a user of SDI to push some data *A* to the data warehouse, and make this data
29 available to all other agents via the SDI query-execution module, and to retain local
30 control over other data *B* on the client machine. In this section we discuss this and other
31 variations, and explain how the functionality of the central SDI data warehouse can be

replicated on a client machine in a *client-side SDI database*, to allow a user to leverage the value in local data when interacting one-on-one with another agent. For example, a consumer's shopping agent might wish to use local data about its preferences to allow an online vendor to configure its services, but without releasing the information to the vendor. We explain how this can be supported within SDI.

We allow the following variations:

1. A user can store some information on his/her local client machine, with that information provided to other agents directly, or retained for querying by other agents on the local client machine as in the system of query-execution on the central SDI database.
2. A user can use a light client machine, perhaps a mobile computing device, which is periodically configured with information stored on the central SDI database that is relevant to the current task. This information can then be used directly by other agents in interactions.
3. A client machine can be automatically configured with information about a user when the user is in its proximity, again to allow direct information exchange with a local agent.
4. The client machine can maintain all information, with virtual links from the central SDI data warehouse to the information. The system operates as before, with information provided if conditions are met, etc.

There are two basic motivations at work. One, a user might prefer to keep control over some information, even keep control away from a central SDI data warehouse, in case it is compromised in some way. We can allow a user to maintain data on a local client machine but seamlessly receive the same benefits as if the data was pushed to the central SDI data warehouse, with virtual links to the client machine. Query-execution can

7.3.1 Data in SDI data warehouse

In Figure 5 we illustrate the situation where agent 1 has its personal information about the user stored in the central SDI data warehouse, and provides agent 2 with the ability to process a query on the data warehouse and access information about the agent. The query might either request information about the agent's profile, or provide a method to execute in the data warehouse, where the result of the method is an optimal action for Agent 2 to make in providing personalized information and/or services to the user with agent 1.

7.3.2 Data on the Client Machine (client-side SDI database) and in SDI data warehouse

Figure 6 shows an alternative variation where data is stored on an agent's client machine and in the SDI data warehouse, and the client-side SDI data warehouse can respond to queries from agent 2 in the same way as the central SDI data warehouse. In particular, as we suggest later in this document, this variation allows an agent representing an on-line consumer to store very personal information on a local machine, and for example only allow other agents to benefit from the *results* of analysis on that information without receiving the actual information.

7.3.3 All information stored Client-side.

Another variation on the basic SDI architecture, is that all information is stored only as *virtual links* within the SDI data warehouse, with physical storage on an agent's local client machine. Preferably, in order to leverage the value of information the client machine would be always networked with the central SDI data warehouse. A current technical solution which seems appropriate would use a *Set Top Box*, connected via cable TV lines to the SDI network. Many households in the US have set top boxes, and this

1 would provide a quite convenient way to leverage the value of information but keep it
2 secure against inappropriate access in a central data warehouse.

3
4 In this variation, to support a personalized session with another agent the data is stored
5 locally, and can be processed using a client-side SDI data warehouse with a local query-
6 execution module. To also support data mining by agents of a shared database, the client-
7 side data records could useful push *data types* and *price rules* to the virtual shared
8 database, to allow more efficient searching. This variation describes a *distributed SDI*
9 *data warehouse*, with indexing performed over on-line client machines.

11 8. Distributed Query Execution

12
13 In addition to allowing distributed data location, the system of SDI allows *distributed*
14 *query execution*, via client-side query-execution modules. The advantage with this
15 variation is that data can be physically stored on trusted local machines close to a user of
16 SDI, and the queries can be physically executed on the machines so that not even
17 intermediate results are available outside of a local firewall protected system. There is no
18 technical solution provided within SDI to prevent agents sharing information to other
19 agents, other than to prevent agents from receiving that information in the first place. We
20 do this using three many techniques: pseudonyms, information randomization (see
21 Section 10), and query-execution in the SDI central server without release of raw data
22 inputs.

23
24 We can implement safe client-side query execution with direct replication of the key
25 functionality of the query-execution and price-module of the central SDI data warehouse
26 on a user's client machine. A querying agent can push the same method to an agent's
27 client machine as it would submit to the central SDI query execution module, and an
28 agent can execute the method locally and release information consistent with his/her data
29 release rules. The SDI client-side agent can be configured, via the profile management
30 and data-release policies, to control the type and amount of profile information released

1 to the central SDI data warehouse. We discuss this in some detail in Section 9, on
2 “distributed data management”.

3
4 The challenge is to be careful that the results of a query do not reveal too much
5 information, but this is solved using the same *price rules* as in the central SDI data
6 warehouse. The mode of interaction supported is QUERY(PID), i.e. execute a query on a
7 particular pseudonym ID, and the client side query execution module ensures that the
8 vendor has a certificate from the client-side SDI agent to query the information in the
9 local database about that pseudonym. The control over the profiles that a vendor can
10 access, coupled with the price rules to ensure that information is perturbed as necessary
11 and that not too much information is released, and to allow an agent to leverage the value
12 of information.

13
14 A user’s agent can retain control over local information by limiting the information that
15 can be released in response to a query. For example, although a query can be complex
16 and access a lot of profile information on a user’s client machine, we suggest an
17 additional protection that controls the information that a vendor can receive for future
18 use. It is possible to limit the response that the vendor receives to a fixed message size,
19 e.g. to 10 bits for example, to provide a very strong overall control on the ability of a
20 vendor to use information again in the future.

21 22 8.1 Example: Customized Vendor Web Pages

23 The method, called *safe client-side query execution*, has a direct application to systems
24 for personalized information delivery, where it is not desirable to release profile
25 information to an information provider, but preferable to allow the information provider
26 to provide methods (e.g. queries) that are executed on a client machine, and personalize
27 information before it is provided to the end user.

28
29 A key application of this technique is in B2C on-line e-commerce where a vendor can
30 push methods for personalization of a virtual shop front to a user, and the user’s SDI
31 enabled client machine can implement the methods (which are queries) on the client, and

4 At present advertising networks such as DoubleClick (www.doubleclick.com) are able to
5 track a user as he/she browses across multiple sites in the DoubleClick network because
6 the ad server can place a “cookie” text string on the hard drive of a user’s client computer
7 which identifies that user as he/she browses. SDI is designed to be used on-line in
8 combination with a method to block cookies, replacing them for example with “safe
9 cookies” [Netscape 96] to still permit stateful interactions during a single session with a
10 vendor, for example “shopping basket” style interfaces.

The SDI-client side proxy agent can replace the role that cookies take in automatic user log-in and password checking, through a Janus/LPWA-style implementation [BGGMM97; BGGMM98; GGMM98]. A user's log-in user name and password can be derived from a user's pseudonymous identity, and computed within a cryptographic framework. Janus also provides a technique to allow a user to receive e-mail pseudonymously.

A core technique in SDI is to use a unique pseudonym for each vendor, and then selectively provide vendors with profile information from across multiple pseudonyms; either anonymously within the central SDI data warehouse for data mining purposes, or alternatively via client-side personalization so that a vendor can leverage a user's wide profile data without receiving *direct* access to that information.

24 In providing profile information to a vendor during an interaction, so that the vendor can
25 provide targeted products and servers to the user based on data mining that it has
26 performed on aggregated user data in the central data warehouse, the user's agent must be
27 careful to protect the identity of a user. This means that the agent must not reveal
28 information under pseudonym P_1 and information under pseudonym P_2 to another
29 vendor that allows the vendors to link the pseudonyms and reason that the agent
30 represents the same user.

1 We describe in Section 10 the technique of random noise perturbation to release
2 information in response to queries, and another simple technique is to carefully protect
3 particularly sensitive information (e.g. social security numbers, etc.)
4

5 8.2 Client-side Query Execution 6

7 As described earlier, a key variation of SDI retains a local data warehouse and query
8 execution module that contains information specific to a single agent, on the client
9 machine of that agent. It is then possible, via the same query execution controls as
10 described in the central SDI data warehouse, to allow vendors to characterize and profile
11 a user based on its local data, but without gaining explicit information about the user's
12 local data. The vendor can push appropriate methods and targeted services. We describe a
13 number of bottom-level applications later in the patent, including personalized web
14 pages, ad-networks, etc.
15

16 We describe one variation in which the vendor provides generic information to the client-
17 side device, that filters that information locally. This is relevant for example in high
18 bandwidth information services such as digital television and satellite systems, where a
19 user's set top box can store program information locally and make local decisions about
20 what programs a user might like to watch and when, without providing any profile
21 information to the head-end server.
22

23 In this variation an agent can submit a query to be executed locally on a *client-side SDI*
24 *data proxy*, that stores information submitted by the local agent. The applications of this
25 technique are exciting, because it allows a user to maintain even more control over
26 profile information but still benefit from personalized interactions with vendors. The
27 client-side SDI data proxy just contains data for the local agent, and allows agents to
28 submit 'personalization queries' of the type 'execute this profiling rule and tell me what
29 product to show the user'. The outcome of the processing might be personalized
30 information, for example a selection of books that a user might be particularly interested

1 in based on its profile information and the methods passed to the agent by the agent that
2 sends general information about the books in its catalogue.

3
4 Figure 7 illustrates the process. The SDI proxy agent maintains a set of profiles for a user,
5 relating to transactions that the agent has performed with other agents in the system, and
6 also other information that relates to the user associated with the agent. A vendor agent
7 can send generic information and a method for personalization, which is executed as a
8 query on the local profile data base, with the results allowing a vendor to decide how to
9 target its products and services to a user. For example, a book retailer might be able to
10 provide a selection of books that a user might be particularly interested in based on its
11 profile information and the methods passed to the agent by the agent that sends general
12 information about the books in its catalogue.

13
14
15 We can provide *rule templates*, which are used to select the type of method that the
16 provider of the information is requesting be applied to the generic information that is
17 passed to the requestor agent. The provider agent only needs to provide *parameters* for
18 the processing. For example, if collaborative filtering is the desired personalization
19 technique then the providing agent provides a list of prototypical cluster centers, but
20 does not need to specify a collaborative filtering algorithm. The interpreter takes the
21 information and methods from the providing agent and selects the appropriate rule
22 template to form a program that is executed as a *safe query* in the CPU, with the usual
23 controls over the amount of data that a query can access.

24
25 The output from this processing is personalized information, that can then be displayed to
26 the user, perhaps to a final test in the control module for applicability.

27 28 **8.2.1 Preventing Information Leakage**

29
30 The challenge is to be careful not to allow the *results* of a query, even those data is not
31 explicitly revealed to an agent that submits a query, to carry sensitive information that a

1 user would rather not reveal. We allow this by: (a) partitioning a user's local profile into
2 a number of different identities, so that a vendor can only access data for a particular
3 profile; (b) allowing only "safe" information to be reported; (c) allowing only "safe" data
4 fields to be accessed.

5
6 One particularly simple technique is to limit the size of a response to a query by the
7 number of *bits*, because this can quite nicely prevent the release of too much information,
8 when coupled with close control over the type of data that is accessed. For example, a
9 user might be willing to allow queries with any amount of "less sensitive" information,
10 with the understanding that the vendor only receives at most 10 bits of information.

11
12 As described above, this profiling might also be a *sequential iterative* process, so that the
13 vendor does not need to transmit a complete decision tree up-front, but request a sequence
14 of information. Only release information to a vendor that we select, and do not allow the
15 same vendor to access too much information over an extended period of time.

16 17 18 8.3 Extended Example: Managing Distributed Personalization

19
20 In this section we describe an extended application of distributed query execution, for a
21 vendor that wants to customize its products and services to a particular user. A central
22 application of SDI is a system where users submit profile information to the shared SDI
23 data base, and receive personalized information and possible cash payments in return.
24 This variation has users optionally submitting information to the central SDI database,
25 and receiving personalized information based on statistical patterns that a vendor is able
26 to assess from the information that is submitted by other agents. A vendor can provide a
27 user with personalized content that reflects his/her interests based on the types of
28 information that other agents with similar profiles like to receive.

29
30 Figure 8 provides a top-level view of one such agent, that we refer to in this section as a
31 *personalization agent*. We consider an agent that maintains an identifier and (partial)

1 profile information for the agents that it has interacted with. Consider, for example, email
2 addresses of customers to Amazon.com. The vendor-side SDI proxy can maintain
3 profiles for each user that it interacts with, indexed on the basis of the log-in identity that
4 a user uses. The agent can also use a "processor" device that takes data about the profile
5 of a user, and the rules that have been collected based on profile information in its local
6 database and from wider SDI query execution, and computes the personalized
7 information that an agent might be most interested to receive.

8
9 The personalization agent's information database (Figure 8) contains all information that
10 the agent can provide to other agents, for example this would be all of today's news
11 stories in the case of an agent representing the New York Times. The rules database
12 contains methods that the agent can use to decide how to select information to present to
13 a user based on profile information about the user. The profile information can be stored
14 local to the personalization agent for the information that the agent has collected about
15 other agents, other information will be stored either in a central SDI database, or client-
16 side at the user's SDI proxy. The communication module receives requests from
17 information from agents, and information about agents' profiles, either from agents
18 directly or from the central SDI server. The communication module also sends rules and
19 profile information to other agents, requests for information to the central SDI server, etc.

20
21 In Figure 9 we look inside the personalization agent's *control module*, which is the
22 module that receives requests for information from the communication module, and
23 checks whether the agent has the type of information that the agent requests, and then
24 recovers the profile for the requesting agent and sends that to the processor. The
25 processor selects a rule from the rule database to use in deciding what data to send to the
26 requesting agent, and can also select a method from the rule database to accompany the
27 data when it is sent to a requesting agent.

28
29
30 The control module receives a request for information from the communications module,
31 and checks with the information database for its capabilities to provide information. A

1 query can also specify no particular information. The next step is to consult the profile
2 database, and investigate whether the agent already has profile information about the
3 user, which it can use in the processor to select a rule that will decide which
4 information to push to the agent.

5
6 The processor returns a rule, and the control module consults the information database,
7 and recovers the required information. Finally, the control module forwards the
8 information, and optionally a method to further process the information at the requesting
9 agent with profile information stored locally to the requesting agent.

10
11 Figure 10 illustrates the role of the *processor module* in the personalization agent. The
12 module receives the request for information from an agent, and any profile information
13 that the personalization agent already has about that agent or that is provided with the
14 information request. The processor consults the rule database, and selects a rule with
15 which to decide what information to send to the agent. The processor can also decide to
16 send a rule with the data, that will be processed on the client of the agent making the
17 request for information based on local profile information to that agent.

18
19
20 The decision module within the processor (Figure 10) selects a suitable rule, and decides
21 what information to provide to the agent. This rule is passed to the *threshold testing*
22 module. This is where a decision is taken about whether further customization should be
23 performed local to the requesting agent, based on its local profile information. If the rule
24 provides sufficiently well personalized information then the processor instructs the
25 control module to process the information with the rule, and send the results back to the
26 information gathering agent. Otherwise, for example in the case that the personalization
27 agent does not have accurate profiling information about the requesting agent, the
28 processor will instruct the control module to partially process the information as best as it
29 can, and then send that processed information and an *additional rule* to the agent for
30 additional processing.

9. Data Management

The invention of SDI, in its most general form, refers to the system for a privacy-protected market for data interchange between multiple parties. In this section we discuss possible methods for an agent to manage its disclosure of data to the data interchange, and to other agents in the system. We allow agents to adopt different identities, and submit different information under different identities. The idea is that a pseudonym prevents other agents from exchanging information. Given an identity management policy, the *data management policy* is then used to determine what data to submit to the central data server, or release to other agents, under a particular pseudonym. The agent needs to choose *what* data to release, and what *price rules* to associate with that data.

We propose client-side “proxy agents” to intermediate the interactions between agents, and manage data provided to the central SDI data warehouse; and methods to manage an agent’s identity in interactions with other agents, for example via pseudonymous and anonymous interactions. The proxy agents control the amount and types of information exchanged between agents. We refer to the proxy as “client-side”, because it resides on the machine local to the agents that participate in the system of SDI. The client-level SDI proxy, implemented as a client program running on the user’s client machine, manages all data release from the user to other agents and to the central SDI data warehouse. The proxy might usefully provide a rule-based interface to allow a user to select appropriate data management policies.

An important application of SDI is to a system for business-to-consumer (B2C) e-commerce, where SDI allows individuals to provide vendors with access to profile information that is collected client-side about an individual as he/she interacts with vendors, but retain control over the amount and level of detail that is made available and collect payment for that information. A key function of the client-level proxy is to control the ability of multiple vendors to track a user, by control of the identity of a user in its interactions and the information released. The client-side proxy agent maintains profile information about a user, but releases that information according to a user-defined policy.

1 The system of SDI, together with these client side proxy agents, allows a user to control
2 the amount of information that is released to vendors and other agents, and allows users
3 to exchange information within a privacy-protected and carefully controlled market place
4 and receive personalized products and services The proxy might usefully provide a rule-
5 based interface to allow a user to select appropriate data management policies. Vendors
6 can run collaborative-filtering type applications on profile data from many online
7 consumers, and use that information to personalize products and services in the future.
8 Vendors can continue to build collaborative filtering models and other personalization
9 models that used extended profiles, and users can receive the benefits of personalization
10 without releasing profile information to a vendor; e.g. by personalization within the
11 central SDI query-execution module or by personalization on a user's client machine.

12 13 9.1 Overview

14
15 The basic assumption is that there is no technical method to prevent vendors from sharing
16 information once they have that information, but it is possible to reduce the value of that
17 information and the privacy implications for a user by making it very difficult to
18 determine that two agents represent the same user in two different interactions.

19
20 The ability to interact pseudonymously with other agents, and adopt a unique (temporary
21 or persistent) pseudonym for each agent-agent interaction, is not in itself novel. This is
22 described in the literature, for example as early as the work of Chaum [Chaum 85], and
23 others more recently [FR 98; Cranor 96; BGGMM 97; BGGMM 98]. What is novel is the
24 central SDI privacy-protected marketplace, that respects metainformation that agents
25 associate with data, to carefully control the availability of data to other agents. Also novel
26 is the method of 'distributed query execution' (Section 8), that allows an agent to push a
27 method to another agent, and take a decision based on private information without
28 learning the information.

29
30 In Section 9.2 we present example identity-management policies. At one extreme, if an
31 agent interacts with every other agent under a unique pseudonym, and never allows two

agents to match its identity based on information that it releases, then the other agents have no way of combining information about the agent. In the business-to-consumer e-commerce example, vendor A cannot share information about consumer Z with vendor B unless vendor B has some way of linking its customer with the customer of vendor A. This can be done at present via cookie mechanisms and advertising networks (e.g. DoubleClick), or simply by matching identifying information (e.g. user name, e-mail address, credit card number) across sites.

In Section 9.3 we describe data-release policies, which are methods that determine what information is released for each pseudonym a user selects, and the rules associated with accessing that information. We also note that an agent can choose to introduce a level of noise-perturbation to data before its release, so that the data is still useful but cannot be used to link an agent across multiple pseudonyms via the data that it releases.

Alternatively, an agent can release data with a reduced level of accuracy, for example within a range of values, and/or restrict the types of data that are released. Methods of data perturbation are discussed in more detail in Section 10.

Finally, note that there are two parts to the data management policy. One relates to the data that a user releases to the central SDI data warehouse under a particular pseudonym, and another relates to the data that an agent releases to another agent in a direct agent-agent interaction. The following bullets summarize the distinction, and note the sections for more details.

- [9.3: Data Management Policy] Manage the Release of Data to the Central SDI data warehouse (with either physical uploading of data, or release of links to physically distributed data); this includes the association of price-rules with data that is provided to SDI.
- [9.4: Agent-Agent Data Management] Manage direct Interactions with other Agents, for example controlling the identities used in interactions, and the amount

1 and type of information released. This is really a special-case of Sections 9.2 and
2 9.3 for data release to the central SDI database.

3
4 In addition to the identity and data management policies, an extension of SDI could allow
5 statements about “fair use” of information, much as is developed in the W3C P3P project
6 [W3C; CR 98; RC 99], that allows a user to state for example whether or not it would
7 like to receive electronic mail in response to information, whether the data can be
8 exchanged with other agents, whether it can be used for personalization. “Fair use”
9 policies are typically protected by “catch and punish” mechanisms, for example with spot
10 checks by special interest groups on business practices, and not with technical solutions.
11 In comparison, we focus on technical solutions to allow a user to control information held
12 by other agents; for example by careful identity management, careful data management,
13 and careful metadata annotations to control queries in the central SDI database.

14 15 9.2 Identity-Management Policy

16
17 Identity management policies allow an agent to assume pseudonymous or anonymous
18 identities in its interaction with other agents, for example in the context of consumer-to-
19 business commerce preventing two vendors from exchanging information about the same
20 agent on the basis of the agent’s revealed identity. An identity-management policy defines
21 an appropriate identity (pseudonymous or anonymous) for different agents with which an
22 agent might interact.

23
24 A pseudonym can be unique for each agent-agent interaction pair, in which case another
25 agent can store information about repeated interactions but cannot exchange that
26 information with other agents on the basis of the pseudonym, because the pseudonym is
27 unique. A stronger form of privacy is provided with an anonymous identity, that does not
28 allow even the same agent to track the agent across multiple interactions. An anonymous
29 identity exists only for a single session—until either party agent changes context and
30 connects to another agent. A pseudonymous identity system can be implemented using
31 standard cryptographic techniques known in the art. For example, the JANUS/LPWA

1 system [BGGMM 97; BGGMM98] allows a client-side proxy to compute a new pseudonym
2 for an agent, and also a new e-mail address and user-name and password for that
3 pseudonym, to allow access to web-pages that require user log-in. In particular, it is possible
4 to compute new pseudonyms for a user without a central register of pseudonyms that could
5 compromise a user's identity, through the technique of "blinded signatures" [Chaum 85].

6
7 The identity-management policy can be configured by a user when he/she first registers with
8 the system of SDI; for example a simple policy might classify agents according to the
9 certificates that they can present, and select a pseudonym according to the agent's class.

10 Other reasonable identity-management policies include:

- 11
- 12 • Absolute Privacy. At the highest level of privacy an agent interacts anonymously
13 with every other agent so that agents cannot learn about the agent across sessions,
14 and cannot personalize future interactions. Furthermore, information about the
15 transaction with the agent, cannot violate a user's privacy when provided to other
16 agents, so long as the agent does not release any identifying information during its
17 transaction.
 - 18
19 • High Privacy. At the next level of privacy, an agent interacts with every agent under
20 a unique persistent pseudonymous identity. This prevents another agent exchanging
21 information about the agent with other agents, unless the agent reveals other
22 identifying information. However, this policy does allow agents to receive
23 personalized information over multiple sessions with the same agent, directly
24 without the vendor using information submitted to the central SDI data warehouse,
25 or stored on a client-side single-user database. Another agent can personalize
26 information, but only on the basis of previous transactions with the agent. In an
27 Internet environment this mode of interaction allows a vendor to track its customer
28 across multiple sessions.
 - 29
30 • Medium Privacy. Use the same persistent pseudonymous profile with groups of
31 other agents, perhaps segmented according to the type of task that an agent is

sell links between profiles, to agents that can present the right characteristics and pay the right price.

Example in Consumer-Business On-line Commerce. Figure 11 illustrates agents 1, 2 and 3; agent 1 maintains one pseudonym for vendors 1 and 2, and a different pseudonym for vendors 3 and 4; agent 2 maintains one pseudonym for vendors 3 and 6; agent 3 maintains a unique pseudonym for each of vendors 1, 3 and 5.

Of course, a vendor might adjust its level of service according to the level of identification provided by an agent. For example, a vendor might provide a more complete service to a user that provides a pseudonym than a user that provides an anonymous identifier, because a pseudonym allows a vendor to collect information over an extended period of time and build a detailed profile that can enable future personalization.

9.3 Data-Release Policy

The client-side SDI proxy provides support for automatic submission of profile data, and other types of data, to the SDI central data warehouse. This includes policies for pricing queries, and policies for introducing random perturbations to data.

There are a number of human/computer interactions that are important in allowing a user to configure these options. . Many of the issues were anticipated in Cranor's work [CR 98; RC 99], in which she describes methods adopted in the W3C P3P (platform for privacy preferences) project, and suggests that users can as a first step select a special interest group with which they are affiliated, use that as a base policy, and then adapt the policy as necessary. Alternative techniques include decision-tree methods as described in pending patent [INSERT US PATENT NUMBER AND NAME], that ask a user a number of questions in order to ascertain an appropriate service. There are many dimensions that a user might like to identify; for example: the type of information that

1 can be released, the types of queries that can be performed on that data, the price that
2 must be paid to perform the queries. For example, we might hard code different price and
3 data-release policy codes, e.g. A B C D..., and allow third-parties to provide maps
4 between data types and an appropriate policy code.

5
6 We view an agent's complete data release policy as an intersection of its data
7 management and identity management policies. As with identity management, a simple
8 variation will define a fixed number of data management policies, stating types of data
9 that can be released and conditions, and a mapping from agent certificates to policies. An
10 agent can receive whatever policy it presents a certificate for.

11 12 **9.3.1 Data Validation**

13
14 The SDI client-side proxy can also provide a useful service of validating data in profiles.
15 This can be important in scenarios in which it is possible that a user can try to misstate
16 information in the database for personal gain. For example, consider a consumer-business
17 e-commerce system in which vendors determine user discounts for products based on
18 their profile information. Clearly all users will attempt to adopt the profile for the
19 cheapest price, if that is known, and if that is possible. We can allow vendors to specify
20 that they will only follow validated information, for example if a user's profile states that
21 he/she is very price sensitive, then it is more useful to know that if the assessment is
22 based on validated transactions performed via the SDI system. In general terms,
23 information is valuable within a marketplace if the information is accurate, and we can
24 use client-side validation to achieve that goal.

25
26 Consider another business-consumer e-commerce example. In general it is useful to allow
27 an agent to duplicate information across different identities in the SDI data warehouse,
28 e.g. my Zip code might well be information that I am very happy to associate with all of
29 my profiles because it is not too revealing of my identity. However, as a vendor I might
30 like to know that when I execute a query I am not paying to receive duplicated data
31 records for the same user. One useful way to prevent this is to allow a user to also state

with some records “this is the only user profile for which I am choosing to submit this piece of information”. For example, if user A purchases book X then the user might choose to associate information about that book purchase with just one of his/her profiles in the SDI data warehouse, e.g. the profile for the “book-reader”. Stating this allows a vendor to specify within its query that it is only interested in receiving information associated with profiles that include this “exclusivity” claim. The SDI client-side proxy can provide a guarantee that the user only associates data with one profile, for example using a cryptographic method, e.g. signing the data record with a key to indicate that it is a unique record.

9.3.2 Data Perturbation

In addition to an agent’s identity management and data-release policy an agent must be careful that it does not compromise its identity-management policy by releasing identifying information. For example, an agent must be careful not to release the same piece of information X under pseudonyms P1 and P2 if it is unlikely that the information would relate to two different agents. In the next section we discuss the concept of data perturbation in some detail. The idea is that an agent must add enough noise to any information that is released to prevent identification by another agent, unless the other agent already knows the agent’s true identity. This date-perturbation module can run on top of identity and data-release methods. Without careful management of the control of information a pseudonymous identity management policy is redundant. Remember, whenever pseudonymity is compromised there is no technical method to prevent vendors and other agents from exchanging information about the agent.

Agents are careful to reveal only information that will not allow a vendor to link the identity of a user across multiple pseudonyms, defeating the identity management policy. Pseudonymity can be broken whenever an agent reveals the same piece of information, X, to multiple agents, e.g. A and B, and that information X has significant discriminative power. Agents A and B might be able to deduce with quite high probability that it is likely that the

1 two pseudonyms refer to the same agent; and therefore combine their profile information
2 and defeat the user's profile management policy.

3

4 Consider an extreme example in business-to-consumer e-commerce: if agent 1 reveals its
5 social security number to vendors 1 and 3, its pseudonymity is broken. Vendors 1 and 3 can
6 now communicate and deduce that the agent that interacts under pseudonym P1 and
7 pseudonym P2 is the same agent, and therefore share profile information that the agent
8 would like to isolate from each agent. The pseudonymity is broken in this case because
9 social security numbers are unique identifiers. A slightly more subtle example is to consider
10 an agent that provides a combination of profile information to two vendors, where the
11 information taken together is suitably revealing that the vendors can conclude that the
12 pseudonyms relate to the same user with high probability.

13

14 At the other extreme, we might simply provide no personal identifying information, such
15 that in a business-consumer e-commerce application the vendor knows only the type of
16 product that the agent is looking for, or the request for information made by the agent. In
17 this situation we can use the methods disclosed in [SECTION XXX, END OF BOTTOM-
18 LEVEL] to complete a transaction with pseudonymous physical mail and pseudonymous
19 payments.

20

21 A slightly more advanced method is to classify information as identifying and non-
22 identifying, where information is non-identifying if no amount of the information can break
23 a user's identity; e.g. information common to many other users, perhaps the city in which a
24 user lives, a rough salary range for a user, a user's banking institution, etc. In comparison,
25 identifying information might include a user's street address, a user's favorite book, a user's
26 exact salary, etc.

27

28 The data-perturbation approach is to select the information to release, via an agent's data-
29 release policy, and then add a necessary amount of random noise to that information to
30 protect a user's identity. We can add a small amount of random noise to data, enough to
31 prevent identification, but not so much that the data has no value to another agent. For

example, my year of birth and Zip code might be almost as valuable for customization purposes as is my full date of birth and street address.

9.3.2 Click stream data.

One type of data that is especially interesting in a browsing environment is click stream data, which is stored at the client machine and represents a sequence of clicks that a user has executed, possibly across multiple vendors. The data can be stored on a user's local client machine and periodically released under an appropriate pseudonym to the central SDI data warehouse. The client-level proxy server that runs on a user's host machine is in a unique position of being able to monitor the user across different pseudonyms and across different vendors' sites. The client proxy might also collect information about:

- The information that is displayed to a user (e.g. the text, the pictures, etc.)
- Information typed at the keyboard, and profile information transferred from the client machine to a vendor.

The data is gathered by passively observing the actions of the user, and not by direct question-and-response. Possible click stream data release policies include the following data-release policies:

- A. Release no information.
- B. Only release data on the URLs of the most recent sites visited.
- C. Release data about the URLs of the most recent sites visited, and the information displayed to the user.
- D. Release data about the URLs, the information displayed, and the information entered by the user.

in combination with a suitable identity-management policy. For example, an agent might state that all click stream data should be released under the pseudonym that the agent adopts for a particular transaction, i.e. all click stream data with amazon.com should only

1 be associated with my “buying books” pseudonym. An alternative policy might state that
2 click stream data should be stored under a unique pseudonym, and not identified with any
3 of a user’s other pseudonyms. The agent might then allow a vendor to perform limited
4 queries on that data, for example only query information related to particular domain
5 names. Many variations are possible: e.g. use a unique pseudonym for each new URL
6 domain, i.e. whenever a user skips to a new site, submit click stream data under a new
7 pseudonym.

8
9 Click stream data can be subject to random perturbation, just like standard data-- for
10 example removing time-stamp information and adding noise to the URLs that a user clicks.

11
12 There may be other click stream data release policies worth considering outside the
13 preferred embodiment discussed above. For example, Intermind’s patent number XXXX,
14 entitled XXXX, provides for the release of a user’s information according to the terms
15 and conditions of the user’s own data disclosure policy.

16
17 Such an arrangement could be further enhanced by conditioning the quality of the data
18 released by the vendor to the user on the strictness of the user’s own data disclosure
19 policy -- users could then be rewarded to share more of their clickthrough data. Various
20 levels of security could also be guaranteed to the user releasing such information – for
21 example the data perturbation technique could be used to ensure that the vendor only
22 receives aggregate information about his visitors.

23 24 9.4 Agent-Agent Data Management

25
26 It is important that an agent that submits data to the central SDI data warehouse also
27 maintains careful control over the data that is directly released to other agents through
28 one-to-one interactions. It is necessary to prevent a “black-market” in data. Furthermore,
29 a user might simply prefer that another agent (for example representing a vendor) does
30 not know certain pieces of information.

1 With respect to preventing a black market in information about a user, it is important to
2 prevent two different agents linking the identity of an agent under two different
3 pseudonyms, based on comparing information provided by an agent under each
4 pseudonym. We do not propose a technical solution to prevent these agents exchanging
5 information (outside of SDI) about the agent when this type of linking is possible. Such
6 violations can not only affect a user's privacy, but might also decrease the value of
7 information provided to the SDI data warehouse, since information can be readily
8 exchanged between agents outside of the constraints of SDI.

9
10 Fundamentally, the policy under which an agent handles data release to another agent is
11 no different from the policy with which data is released to the central SDI data exchange.
12 This is a special case, in which the agent knows that it is providing explicit permission
13 for an agent (specifically the agent with which it interacts) to receive all information that
14 it releases. As indicate above, the pseudonym and the data release to another agent can be
15 determined by certificates that the agent is able to present, and an agent's local data and
16 identity management policies.

17
18 For example, suppose that the policy defines that certificate A confers the right to receive
19 a persistent pseudonym, and also receive links to the pseudonyms in set P1. Suppose the
20 certificate B confers the right to receive a anonymous pseudonym, and receive no links,
21 and certificate C confers the right to receive a persistent pseudonym, and links to
22 pseudonyms in set P2. Finally, suppose an anonymous profile is the default. Now,
23 suppose a vendor presents certificates A and C. This vendor receives a persistent
24 pseudonym, and links to pseudonyms in the union of sets P1 and P2; a vendor that
25 presents A and B receives a persistent pseudonym and links to pseudonyms in set P1; and
26 a vendor that presents no certificates receives an anonymous profile and no links to other
27 pseudonyms.

28
29 9.5 Light Clients: Web-Centric Data Management

30

1 In this section we describe a variation that pushes data management to a centralized web-
2 server, so that "light"
3 mobile computing devices can achieve the same functionality as a more powerful desk-
4 top computer. The convergence of telecommunication and computation technologies is
5 leading to many new mobile devices, but they often have limited computational
6 resources--- processing power, memory, storage space, battery life-time. It is an
7 interesting problem to provide these mobile devices with general computational abilities,
8 for example leverage centralized computation to do such compute-intensive tasks as data
9 mining and information extraction.

10
11 We explain how the functionality of the client-side proxy agent, for example in managing
12 the identity and profile information of a user as he/she interacts with other agents in a
13 networked environment, can be pushed to a central server in the system of Secure Data
14 Interchange. This is useful in the case of a light client with limited data storage and
15 information processing capabilities. This enables a user to maintain a persistent SDI profile
16 across different client machines, for example at work and at home. The profiles and
17 pseudonyms that an agent uses to interact with other agents can be stored in a central
18 computational server, with remote user access. Alternative technologies, such as smart-card
19 technologies, might allow information about an agent to be physically stored with a mobile
20 agent, but use central servers and local compute devices to perform computation.

21
22 The first step is to support user-identification, for example via a mobile key-card [Chaum
23 85; Chaum 91], password and question-response, a voice-test, an iris-scan test, etc. With
24 this, the client's proxy agent can authenticate its identity to the SDI profile management
25 server, and request that the profile manager initialize a personalized interaction session
26 with another agent. Consider for example a physical SDI device in a store, that requests
27 notification whenever agents with light SDI clients are in its physical proximity. At this
28 stage the light client and initiate an exchange of information between the SDI store agent
29 and the SDI central servers, and receive the benefits of a personalized interaction.

30

Figure 12 illustrates a sequence of steps that can allow a light client to execute a customized interaction with another agent. Suppose agent 1 is a light client. Initially agent 2 contacts agent 1, requesting profile information so that agent 2 can provide a customized service to agent 1. Agent 1 receives the request, and contacts the SDI profile management server, which implements the functionality of an agent's data management and identity management policies, and provides an identifier for the appropriate agent pseudonym to agent 2. Agent 2 can then request information about agent 1 directly from the central SDI server, and finally provide personalized information and services to agent 1.

The pending patent application "A System for Location Enhanced ..." [FILL IN NAME AND PATENT #] describes a method to identify a user based on location, which can allow automatic detection of close SDI-enabled agents based on physical proximity. The methods taught in the LEIA patent also suggest the use of a common user identifier, which could via simple look-up act as the Unique User identifier in the central SDI database. Other biometric variations, e.g. via Iris scan technology promise to allow personalized interactions between people and other agents that they approach in the physical world, with sessions automatically configured based on identification and principles of consent. Location based filtering can be further used to allow relevant information to be delivered to a user based upon his/her physical location. It is also possible to pre-cache personalized information directly to the user's client (and server) thus overcoming the bandwidth bottleneck or wireless connections.

Web-centric profile management can be used in an environment of mobile users that interact with many different computational devices, for example ATM machines, point-of-sale terminals, etc. A simple variation of SDI allows a user to allow a machine with which he/she interacts to be automatically configured for a personal session, based on carefully controlled queries from that machine to the central SDI data warehouse. A common "user identifier" be it cards, codes or biometrics can be used to identify and personalize the local SDI client machine. The user may also be identified during off-line

1 transaction via his/her credit or debit cards, for example on ATM machines and POS
2 kiosks, providing promotional offers and coupons.

4 9.6 Smart Cookies

6 Current practice in web-browser based consumer-to-business electronic commerce is to
7 use cookies, which are identifiers placed on a user's hard drive, to identify a user across
8 an extended period of time. For example, if I access the New York Times from my home
9 personal computer on Monday, and then again on Tuesday, the New York Times server
10 can identify that I am the same individual and build a profile of my interests, i.e. the new
11 stories which I choose to receive first each day. This type of information about all of the
12 users that read the Times can allow collaborative-filtering type techniques and
13 personalization of information in the future, such that my "front page" is different from
14 the front page of someone else.

16 However, cookies have the unfortunate side effect of allowing an individual to be tracked
17 across the web pages of different vendors, for example across the web page of the New
18 York Times (www.newyorktimes.com) and Amazon (www.amazon.com), if the Times
19 and Amazon both embed content from the same third party in their pages. This happens,
20 for example, with the DoubleClick advertising network. DoubleClick
21 (www.doubleclick.com) operate a virtual network of pages, and can track a user across
22 any page within their network, and gather a very comprehensive user profile. Although
23 the cookie mechanism is designed so that only vendors with the same domain name can
24 access cookies on a user's hard drive, they can easily be used to profile users across
25 multiple vendors, for example with a double-click style network that embeds a universal
26 advert server within each page.

28 The system of SDI allows controlled personalization, such that a proxy-automated log-in
29 session where the proxy presents a user's pseudonym to a vendor allows that vendor to
30 track a user over time at its own web site, but a vendor cannot track a user across web
31 sites. The SDI client-side proxy agent will disable cookies in their current form (although

BGGMM98;GGMM98] or W3C proposals [CR98; RC99; W3C-OPS 97]; essentially the pseudonym and an associated password that a user adopts for a vendor allows log-in to that vendor.

Also important when supporting a system that allows a user to browse pseudonymously is that the physical attributes of a network system are removed of their identifying characteristics. For example, another role of client-side data management agents is to strip the 'from' field in a HTTP/TCP message.

9.7 Implementation Details

In an Internet browser environment the client-side proxy agent that provides distributed data management for an agent might be implemented as a plug-in into the browser, that can for example be downloaded from a central SDI server. The browser is then configured to use the SDI proxy as its proxy, and the SDI proxy itself connects through a user's ISP (or other intranet gateway) to the Internet, and on to other vendors.

The user provides his/her SDI proxy with personal information, such as his/her name, mailing address, and e-mail address. The client-level proxy registers then registers the user with the central SDI server, providing the server with the name, address and e-mail address of the user. Other basic user information might include demographic information, for example a users job, marital status etc. The client proceeds to automatically generate a unique SDI user ID code, and a private key to allow future authentication of its log-in.

The client can create a unique public key/private key pair. This key pair can be generated only once for a person, and although the central SDI user ID server does not know the key pair, the server can verify that a key pair is only generated once-- because a new user must present proof of identity to establish an account. The client generates a unique user identifier, UUID, for example with the methods taught in [Chaum 85; Schneier 92]. The UUID can then be blinded and signed to certify that a user is registered with SDI, using Chaum's technique of blinded signatures so that the certifying agent does not the identity of

1 the UUID that it signs [Chaum 85, Chaum 92]. The client-level proxy can remove the
2 blinding factor, to compute a signed UUID that it uses when it is necessary to generate new
3 pseudonyms and request new certificates.

4
5 The client-level proxy can now sign messages with its private-key, and provide the signed to
6 UUID, to verify that (1) the UUID represents a validated user; (2) it is the client-level proxy
7 authorized to act for the user, because it has the private-key associated with the UUID. The
8 client-level SDI proxy uses the private key to authenticate messages that it sends to other
9 modules within SDI, such as Pseudonym administering servers. The unique user ID for a
10 user does not carry any information about the user, its sole purpose is to provide a unique
11 identity. The unique UUID can also be used to generate new pseudonyms for users that are
12 certified one-time for a particular vendor, so that the user certifies that he/she maintains the
13 same profile for all interactions with a vendor.

14
15 At this stage the central SDI server might verify the identity of the user, and also check that
16 the user is not already registered with SDI. The method for verifying the identity of a user
17 could include requesting that the user provides his/her social security number, or some other
18 institutional solution that is used for this purpose. In the future we could envisage an
19 electronic system for such an identity procedure, but the method might require for the user
20 to execute this initial step in person with the presentation of a recognized photo ID. The
21 central SDI user ID server maintains a database of all users that are registered with SDI, and
22 checks that that the user is not already registered with the system of secure data interchange.
23 The central SDI proxy may also provides the user with a signed certificate of some universal
24 identifier, such as its Social Security Number, that the user can use to generate other
25 certificates from certifying agencies, to be able to gain certificates under pseudonymous
26 identities.

27 28 **9.7.1 Initializing Identity and Data Management Policies**

29
30 The client-side SDI proxy might provide a rule-based interface to allow a user to select
31 from a menu of defaults an appropriate profile management policy, or to configure with

the aid of a decision tree an appropriate profile management policy for a user. Profile management policies define how a user will interact with various classes of vendors (depending on the nature of the business that the vendor is engaged in), the kinds of uses to which the transactional information that a vendor collects can be put to, and the amount of information that a vendor which collects profile information about a user is authorized to release. The client-level proxy manages a user's interactions with vendors, to keep them within desired policies.

9.7.2 Generating a New Pseudonym

The method of blinded signatures [Chaum 85; Chaum 92] provides a useful technique to generate new pseudonyms for a user, without any centralized database that stores pseudonyms. The client-level SDI proxy can generate a new identifier, comprised of a sequence of bits, that will be unique with a high degree of probability. The identifier is "blinded" and then submitted to a trusted-third party to be authenticated for use as a pseudonymous identifier. When authenticated the blinding factor is removed, and the final signed pseudonymous ID can be used as a new identifier. The signature can allow a user to associate certificates with that identifier, and also to verify that the user has only a single pseudonymous ID with a particular vendor.

Pseudonym administering authorities (PAS) cannot build dossiers of the pseudonyms, because users submit "blinded" identifiers. The only information that a PAS has is the list of unique vendors that a particular user has registered with. The pseudonym administering server can be operated by an agent with a trusted relationship with a particular vendor.

Every SDI user has a unique identifier, a UUID, that is presented to the PAS with the new pseudonymous identifier to be validated. The PAS can verify that this is the first pseudonym for a particular vendor. Each Pseudonym administering server has a public key / private key pair (PKPAS, SKPAS) for each Vendor for which it validates new pseudonyms. A signed pseudonymous identifier, signed with the private key of PAS, verifies that the pseudonym is valid. Cryptographic techniques ensure that the signature cannot be falsified, and allow

1 another party to verify the signature with the public key of the PAS. This enables a vendor
2 that receives the pseudonym to validate that the pseudonym is unique for the user, to enable
3 persistent interactions across multiple sessions.
4

5 An agent can use its pseudonymous identity to initialize transactions with a vendor, and
6 also to encrypt messages. However, it may be more efficient to exchange future messages
7 with a shared key pair—this is more efficient to implement than an asynchronous key pair
8 cryptographic solution. Messages can be encrypted with the shared key, that only the user
9 and the vendor know. This (1) validates that the message is from the sender; (2) ensures
10 that only the intended recipient can read the message.
11

12 **9.7.3 Automatic Client-Side Profile Management**

13
14 When a user clicks to a new URL the SDI proxy agent must determine (a) the identity to
15 adopt with the vendor; and (b) the profile-management policy to adopt with the vendor.
16 The first-step can be performed by local look-up, in a stored table of the vendors that the
17 user has interacted with. Should the vendor be a new vendor, then it is necessary to
18 implement the identity-management policy based on certificates provided by the vendor,
19 and then either use an existing pseudonym or alternatively adopt a new pseudonym.
20

21 With an appropriate pseudonymous identifier the proxy agent can connect to the vendor,
22 sending the vendor its signed identifier to identify itself. The proxy server continues by
23 sending the data privacy policy for the user with this vendor, signed with the private key
24 for the pseudonym. This serves three main purposes: (1) it demonstrates to the vendor
25 that the user is indeed the owner of the public key PKP, because the user has the secret
26 key that is associated with it; (2) it informs the vendor about the data-privacy policies that
27 the user requires; (3) whenever the vendor submits information about transactions with
28 this user to the central SDI server it must also submit this certificate to verify that it is
29 following the user's guidelines.
30

1 The client-side SDI proxy agent might also provide a “connection-certificate” to provide
2 the vendor with certain abilities to access information about the user in the central SDI
3 data warehouse, as described in the earlier section. When a user connects to a site and
4 provides a certified public key, the first-level proxy server also provides a time-stamped
5 certificate of connection, $S((PK*V, T), SKP)$, where T is the current time, $PK*V$ is the
6 unique public key of the vendor, and SKP is the secret key of the user for the pseudonym
7 that it uses with the vendor. This “connection certificate” is used by the vendor to request
8 a profile-release from the central SDI server.

9
10 The proxy agent then implements the profile-management policy, as it relates to the
11 certificates presented by the vendor. First, the agent might submit the basic profile
12 information to the vendor, for example a user’s age, nationality, state, sex—anything a
13 user is happy for any vendor to know. This is the basic profile that is configured by the
14 user during initial registration with SDI. Later, the vendor might request profile
15 information, that can be provided if the vendor is authorized to receive such information.

16 17 **9.7.4 Leveraging Existing Standards**

18
19 The architectural framework outline above can be implemented with a number of existing
20 technical methods. One approach is to use the Extensible Markup Language (W3C-XML) to
21 encode information exchanged between client agents and vendor agents, for example profile
22 information and requests for information. This has been suggested by the W3C consortium.
23 The eXtensible Markup Language (XML) proposal of the Worldwide Web Consortium
24 working group on SGML provides an ideal standard for representing such information.
25 XML allows meta-content to be included with documents, machine-readable information
26 that enables documents to be processed by client software. Augmenting web documents
27 with structured information enables clients to respond to requests for profile information.
28 XML can represent rich data structures, and that allows a grammar to be defined for
29 information that allows data to be automatically verified for correctness.

30
31 Example: An XML Representation for a Profile Request Message

1

2 A vendor might request profile information from a client-side SDI proxy agent. A vendor
3 sends a 'Request Profile' message, and the client-side SDI proxy agent responds with a
4 'Profile Update' message, that contains profile information, in accordance with a user's
5 profile management policy. The request-response mechanism can be implemented using the
6 standard HTTP Post/Response mechanism in conjunction with XML message types. The
7 'Request Profile' message can be represented in XML as:

8

```
9      <?XML version = "1.0"?>  
10     <?xml:namespace ns = "http://www.sdi.com" prefix = "SDI"  
11     ?>  
12     <!--doc-->  
13     <SDI:Request> http://www.some_vendor.com </SDI:Request>
```

14

15 and the SDI proxy agent's 'Profile Response' message can be represented in XML as:

16

```
17     <?XML version = "1.0"?>  
18     <?xml:namespace ns = "http://www.sdi.com/clientX12345"  
19     prefix = "SDI" ?>  
20     <!--doc-->  
21     <SDI:Update>  
22         <SDI:Profile>  
23             <SDI:Item> (1231, 0.453) </SDI:Item>  
24             <SDI:Item> (1041, 0.034) </SDI:Item>  
25         </SDI:Profile>  
26     </SDI:Update>
```

27

28 An illustrative Document Data Type (DTD) for an SDI:Profile element type is presented in
29 the next section. The XML messages are included in the body of standard HTTP
30 Post/Response messages. We limit the performance degradation caused by out-of-date
31 profile information that is stored within web pages of on-line vendors by associating "out-

of-date” time stamps with the profiles that are provided by the central SDI server. This mechanism is similar to the “expiration time” tag of a Netscape Cookie message. The frequency with which profile updates need to occur will depend on the speed with which profile information changes. The “out-of-date” time stamp can be included as an additional element in an SDI:Update message.

The system as outlined above can be implemented within the current HyperText Transfer Protocol (HTTP), as a sequence of challenge/response pairs between clients and servers. The HTTP Post/Response mechanism allows clients and servers to exchange data, and this data can be an instance of an XML Document Type, within the body of a HTTP message. The HTTP protocol is the underlying mechanism, with SDI messages contained in the body of the HTTP Post and HTTP Response as XML documents.

Example: A Possible XML Representation of a User profile

The World Wide Web Consortium (W3C) SGML working group developed XML (extensible markup language) to provide an open and extensible grammar for structured data [XML]. An XML document has an associated schema definition to enable an XML-enabled browser to validate the structure of XML data automatically. A Schema in XML is called a Document Type Definition (DTD), and defines the names of tags, their structure, and their content model. XML allows the DTD for an XML file to be identified through a Universal Resource Indicator [URI] in the header of the file (see below). XML also allows URIs for mobile code resources to be referenced, in order to enable a client to process embedded XML data. An XML document must be well formed, and in order to be well formed the tags must form a tree structure. In addition, the DTD allows the structure of an XML document (an instance) to be validated against a particular schema. Senders and receivers must only send valid SDI files. Each SDI message is a valid XML document.

We provide an example XML instance and part of a Document Type Definition for use within the system of SDI. We assume in this example that profile information is represented

1 as a list of attribute-value pairs within an XML document. An attribute is defined by a
2 numeric code, and the value defines the weight of the attribute. For example:

3
4 <?XML version = "1.0"?>
5 <?xml:namespace ns = "http://www.w3.org/OPS/OPS" prefix =
6 "OPS" ?>
7 <?xml:namespace ns = "http://www.sdi.com" prefix = "SDI"
8 ?>
9 <!doc>
10 <SDI:ProfileData>
11 <SDI:Location>
12 <SDI:Geocode> 12321561 </SDI:Geocode>
13 <SDI:DigiMap> http://www.digimap/?12321561
14 </SDI:DigiMap>
15 <OPS:Zip> 19103 <SDI:/Zip>
16 <SDI:/Location>
17 <OPS:Demographic>
18 <OPS:Gender> F </OPS:Gender>
19 <OPS:Age> 26 </OPS:Age>
20 <OPS:Income> 50000-75000 </OPS:Income>
21 </OPS:Demographic>
22 <SDI:ID>
23 <SDI:Pseudonym> P12543 </SDI:Pseudonym>
24 <SDI:PublicKey> 12453246129421 </SDI:PublicKey>
25 </SDI:ID>
26 <SDI:Profile>
27 <SDI:Profile-item> (1242, 0.546) </SDI:Profile-item>
28 <SDI:Profile-item> (56, 0.045) </SDI:Profile-item>
29 </SDI:Profile>
30 </SDI:ProfileData>
31

1 The Document Type Definitions for this document are specified in the header, and include
2 URIs to a DTD of the Open Profiling Proposal of the W3C, and also a DTD of the Secure
3 Data Interchange. The OPS DTD is used to boot strap the SDI DTD, providing tags for
4 common profile information, such as 'Gender', 'Age', 'Income', etc. The section of the SDI
5 Document Type Definition that is used in the above XML fragment is presented below. It
6 makes reference to tags defined in the OPS DTD, and the RDF (Resource Description
7 Framework), a W3C proposal to standardize the structure of Digital Type Definitions for
8 XML documents. XML Name spaces [NS] provide a method for unambiguously
9 identifying the semantics and conventions governing the particular use of property-types by
10 uniquely identifying the governing authority of the vocabulary, for example OPS and SDI in
11 the example above. The URI for a schema can contain a human and machine-readable
12 description of an XML schema.

13
14 <!ELEMENT SDI:ProfileData (SDI:Location?,
15 OPS:Demographic?, SDI:ID?, SDI:Profile?) >
16 <!ELEMENT SDI:Location (SDI:Geocode?, SDI:DigiMap,
17 OPS:Zip?, OPS:Address?) >
18 <!ELEMENT SDI:ID (OPS:Name?, SDI:PublicKey?,
19 SDI:Pseudonym?) >
20 <!ELEMENT Profile RDF:list<SDI:Profile-item> >
21 <!ELEMENT SDI:Geocode #PCDATA >
22 <!ELEMENT SDI:Digimap #URI >
23 <!ELEMENT SDI:PubicKey #PCDATA >
24 <!ELEMENT SDI:Pseudonym #PCDATA >
25 <!ELEMENT SDI:Profile-item (SDI:Attribute-ID,
26 SDI:Attribute-value) >
27 <!ELEMENT SDI:Attribute-ID #PCDATA >
28 <!ELEMENT SDI:Attribte-value #PCDATA >

29
30 The tag '#PCDATA' is used here to represent numeric or textual information, '#URI'
31 declares that an instance of element 'SDI:Digimap' must be a valid URI pointer.

1 guarantee that no malicious party, even for example an employee of SDI with privileged
2 access to the SDI data warehouse, can benefit from highly sensitive information. Second,
3 the data repository in the central SDI data warehouse can be programmed via price-rules
4 for access to data records to allow randomized information to be released in response to
5 queries, perhaps for a cheaper price.

6
7 The method of random data perturbation allows an agent to protect the pseudonymity of
8 profile information which is released within the system of Secure Data Interchange in
9 response to queries executed on the central SDI data warehouse, and also with respect to
10 information that is provided to other agents (e.g. from consumer client machines to vendors'
11 servers in e-commerce applications) during direct interactions. Information is either adjusted
12 slightly, through the addition of small amounts of noise, or rounded or binned in some
13 way—to protect the identity of an agent if it is also releasing the same information under
14 other pseudonyms. The goal in the method of data perturbation is to only adjust data by as
15 much as necessary to protect the privacy of an agent, so that the value of data for
16 personalization and user-profiling is retained.

17
18 As an example, suppose a user lists its last 10 compact disk purchases, their prices, and the
19 dates of purchases to two different on-line music vendors. Suppose the agent uses different
20 pseudonyms with each vendor. Given information about the distribution of CD purchases
21 over the population of agents, the vendors can compute the probability that the pseudonyms
22 relate to the same agent. If it is quite likely that it is the same agent, the vendors can now
23 exchange information about that user, and even collude on price. Clearly, the more revealing
24 the information (e.g. providing more information, location of purchase, last 4 digits of credit
25 card, etc..), the greater the chance of identifying a user.

26
27 However, the system of SDI is built around the concept of gaining value from the release of
28 information, and maintaining careful control over the particular information available to
29 other agents. The method of random perturbation allows information to be released without
30 losing control over the ability of another agent to link the information with other
31 pseudonyms that an agent might use.

1 The main focus here is on random data perturbation, or rounding etc. as a method to
2 prevent the identification of an agent and the linking of pseudonyms. This is critical to
3 the successful operation of a system of pseudonymous data exchange, such as that
4 implemented within SDI.

5
6 Figure 13 illustrates how an agent's price rule can usefully contain an additive cost to
7 reflect an agent's preferences to protect its protection level during query execution. For
8 example, highly sensitive information can have a high price, while other information can
9 have a low price. This additive cost to account for an agent's protection against the
10 linking of its pseudonyms can be universal across all data records.

11
12 We propose a technique to allow a user to control information released and secure a
13 certain protection level, which is an indication of the probability that its true identity can
14 be recognized from information that it releases by an adversary with a perfect copy of the
15 agent's true profile. The randomization method adds a level of noise to information to
16 provide protection, based on the amount of information that is released and the
17 information that is present in a database because of data records submitted by other
18 agents. This is illustrated in Figure 14, which shows qualitatively that if there are many
19 data records then the agent needs to add less noise to achieve the same level of protection
20 as with less data records.

21
22 The method of random data perturbation adds noise to data records to prevent an
23 adversary with accurate information about the
24 values of certain fields in the data record under another pseudonym from linking the
25 pseudonyms, and reasoning that the two users are the same. Alternatively, we can round
26 or bid data values, this has a similar effect.

27
28 Basically, we add enough random noise to make the record sufficiently indistinguishable
29 from other data records. There are two modes of protection, that depend on the amount of
30 information known by the adversary:

31

Strong protection is only possible if there are enough data records from other agents to allow agent **A_1** to add noise to its own data and make it sufficiently like that one of the other data records is the data of **A_1**; in particular the other data records must also be submitted with random noise perturbation, or rounded, such that it is possible that other data records are submitted by agent **A_1**.

12 B. **[Weak Protection]** Assume that the adversary does not know that the agent
13 **A_1** with data records **P** has definitely submitted data into the database.

Weak protection is easier to achieve. It is only necessary to add enough random perturbation to data to make the number of possible data records over the population of possible data records that are supported with the perturbed data record large enough to prevent agent identification.

Randomized data is still useful within SDI for data mining and other applications, so long as the amount of noise which is added to records is small in comparison to the value of a record. For example, we can still perform correlation across fields with randomization so long as the randomization does not destroy any trends between fields. Randomized data is marked as such within SDI, and labeled with the degree of degradation, so that SDI can be aware of the number of records to get relevant accuracy levels, and can report accuracy to customers. There is a tradeoff between the level of privacy protection and the level of aggregation at which responses to queries become accurate. It is possible to add random noise to data but still allow data that is aggregated across multiple records to be quite accurate, so that useful data mining can be performed. Binning or rounding of data does not have the same effect. For example, assuming additive noise and additive aggregation, then randomized data from a number of agents can be aggregated to obtain an aggregate value

1 (e.g. mean) quite accurately as the number of agents increases for noise perturbation with
2 zero mean.

3
4 Random perturbation of individual data records can provide a cheaper and more secure
5 alternative to cryptographic techniques, such as secure function evaluation, for providing
6 information without compromising privacy [Schneier 92]. Randomized data is secure to
7 computational attacks and the loss or theft of private keys-because we degrade the data, and
8 make access to any one data item virtually useless. Cryptographic techniques known in the
9 art provide methods to compute aggregates or other values from encrypted information
10 without first decrypting this information. However, the general-purpose nature of these
11 methods makes them unnecessarily cumbersome for complex problems. In particular, the
12 communication and computation requirements of these methods when applied to the
13 problem of aggregation result in an unacceptable overhead on the system.

14
15 Finally, all cryptographic techniques are subject to compromise by successful attacks on the
16 cryptographic scheme or by the loss/theft of private keys. Such problems are present in all
17 uses of cryptography. In comparison, the method of random perturbation is an information-
18 theoretic technique, and not subject to the same problems.

19 20 10.1 Technical Details

21
22 We propose one possible method to select an appropriate amount of noise to add to data.
23 The noise can be added in two places:

- 24
25 1) At the agent's client machine as data is first provided to the central SDI data
26 warehouse
27 2) In the SDI data warehouse as queries are executed, with the SDI query execution
28 engine implementing data perturbation for agents.
29

30 With numeric continuous data we can select an additive noise term, for example sampled for
31 example from a Normal distribution. The amount of noise is computed to provide an

1 appropriate degree of confidence that an adversary with exact information about its
2 attributes cannot identify the individual. For example, in a Normal distribution, the standard
3 deviation that is required depends on the number of agents that have provided similar
4 information, and how sparse the information is in its domain. The standard deviation is
5 “tuned” to provide the appropriate level of privacy. We do not require that all data is
6 perturbed from the same noise distributions.

7

8 The amount of noise perturbation to add to a data record can be computed using an
9 adversary argument, where we assume that an adversary knows the true values of an agent’s
10 data, and may also (strong) know that the agent has submitted data to the database. We
11 allow an agent to submit or release data with a random noise perturbation, and also can
12 provide the service in the central SDI data warehouse in combination with price rules, to
13 control the accuracy of data released to agents in responding to queries. An agent with true
14 data \mathbf{X} computes perturbed data \mathbf{X}' with additive noise computed from distribution \mathbf{f} , and
15 submits the distribution \mathbf{f} and the perturbed value \mathbf{X}' .

16

17 The problem is more difficult with discrete data, but we can instead select a new discrete
18 element at random from a close neighborhood of the original data point [Section 10.4]. In
19 addition, with multi-attribute data, different, independently chosen noise terms are used for
20 each field of a record that needs to be perturbed by the addition of noise [Section 10.5].

21

22 **10.1.1 Single Attribute and Strong Protection**

23

24 Agent \mathbf{A}_1 has data \mathbf{X} to submit to a shared database, e.g. the data warehouse in SDI.
25 Alternatively, we can assume that the agent is about to send the data directly to another
26 agent. Suppose in this example that \mathbf{X} is a continuous real number.

27

28 Suppose that an adversary knows that agent \mathbf{A}_1 will submit data record \mathbf{X} , and the goal of
29 the agent is to submit data \mathbf{X} with enough random noise perturbation to provide some
30 protection against identification. The level of noise that the agent needs to add to the record
31 depends on the other data records already present in the database.

1 **Pr(true-value = X)** This is computed as the sum of **Pr(A_i) x Pr(X | A_i)** over all data
2 records, and represents the probability that any one of the data records was represents data
3 with true value **X**.

4

5 Agent **A₁** can now select parameters for distribution **f** based on this analysis, so that the
6 probability that it is the agent to submit the new data record is less than **eps**, its desired
7 protection level.

8

9 The rule has the right behavior—the more data present in the system then the more accuracy
10 an agent can use to submit its own information, for the same privacy level epsilon. Notice
11 that the agent is more protected as:

12

- 13 1. **Pr(A₁)** decreases, i.e. with more data records in the database
- 14 2. **Pr(true-value = X / X', f)** decreases, i.e. with more noise perturbation **f**
- 15 3. **Pr(true value = X)** increases, i.e. there are more data points from other agents that
16 might have true value **X**.

17

18 The parameters for **f**, for example the mean and standard deviation in the case of noise
19 generated from a Normal distribution, can be selected to set **Pr(X' | true-value = X, f) = eps**
20 for the case that **X' = X**, i.e. when the random noise is zero. This presents the worst-case,
21 assuming that distribution **f** places the most probability on zero noise.

22

23 Finally, to compute parameters for **f** the agent requires aggregate information about the data
24 in the database that has been submitted by other agents. In particular, the agent needs **Pr(X)**,
25 the probability that any of the current data records could correspond to data with true value
26 **X**. In some cases it is important to receive this information without revealing true value **X**.

27

28 In a simple case, for example when a trusted intermediary such as SDI stores information
29 provided by agents, the agent can simply poll SDI for **Pr(X)** before selecting a level of noise
30 to use to perturb its information. Alternatively, when SDI is computing the level of noise
31 perturbation to add dynamically as information is released in response to queries then the

1 system of SDI can compute $\Pr(\mathbf{X})$ directly. In a more complex case, consider a problem
2 where an agent is releasing information directly to an adversary, and the relevant set of data
3 points are data that the adversary agent has already collected from other agents. In this case
4 the provision of information $\Pr(\mathbf{X})$ must be done within a secret-protocol where the
5 adversary does not learn the value of \mathbf{X} in the process. A straightforward way to achieve this
6 is for the “adversary” agent to post aggregated information about the probability that a data
7 record in its population has true value \mathbf{X} , for a range of different values of \mathbf{X} , and allow an
8 agent to use anonymous look-up in a table.

10 In summary, the following procedure can be used to select an appropriate level of noise to
11 add to a data point:

- 13 a. Choose a distribution family (e.g. Normal, Uniform, etc.), and let \mathbf{Par} denote the
14 parameters that define a specific distribution.
- 15 b. Choose a level of privacy protection, ϵ where $0 < \epsilon < 1$.
- 16 c. Request $\Pr(\mathbf{X})$ from the database, i.e. the current probability that a perturbed (or
17 otherwise) data record in the database has true value \mathbf{X} .
- 18 d. Compute parameters for the distribution to set $\epsilon = \max_{\mathbf{X}'} \Pr(\mathbf{A}_1 | \mathbf{X}', \mathbf{X}, f)$
19 where \mathbf{X}' is the value generated from f .

21 10.1.2 Examples: Strong Protection

23 (a) Uniform Additive Noise Perturbation Distributions

25 Assume that every agent submits information from a uniform noise distribution, centered
26 around its true value. The decision variable in choosing a level of data perturbation when
27 submitting new information is the range of the uniform distribution.

29 Suppose agent 1 submits data point $\mathbf{g}_1 = [4, 6]$, to denote that its perturbed value is \mathbf{X}' , and
30 the value was computed with additive uniform noise $U(-1, 1)$. Agent 2 submits data point

1 $g_2 = [4,6]$, agent 3 submits $g_3 = [3,5]$, agent 4 submits $g_4 = [3,6]$, and agent 5 submits
2 $g_5 = [2,4]$.

3

4 Suppose that an adversary knows that user John has true value $X = 2.5$, and that an agent for
5 John has submitted a data record.

6

7 In this case the adversary can be sure that agent 5 represents John, because $\Pr(A_5 | x = 2.5)$
8 $= \Pr(A_5) \Pr(x = 2.5 | [2,4]) / \Pr(x = 2.5) = 0.2 * (1/2) / (0.2 * 1/2) = 1$.

9

10 In comparison, for an adversary that knows that Mary has true value $X = 5$, then $\Pr(A_1 | x$
11 $= 5) = 0.2 * (1/2) / (0.2 * 1/2 + 0.2 * 1/2 + 0.2 * 1/3) = 0.375$, and similarly for agent A_2 . The
12 probability that agent A_4 is Mary is 0.25.

13

14 Now, a new user, Bill, with agent A_6 wants to provide information about its data $X = 5$,
15 and wants to be sure that an adversary cannot determine its identity with probability greater
16 than $\epsilon = 0.1$. Let d_6 denote the uncertainty selected by agent A_6 , and compute an
17 optimal d_6 as so that $\Pr(A_6 | x = 5)$, i.e. the probability that agent A_6 has true value $x =$
18 5. This is computed as $\Pr(A_6 | x = 5) = 1/6 * (1/d_6) / (1/6 * 1/2 + 1/6 * 1/2 + 1/6 * 1/3 +$
19 $1/6 * 1/d_6) = 1 / (4/3 * d_6 + 1)$, and is less than ϵ for $d_6 > 6.75$. Therefore agent A_6
20 should generate a perturbed value with additive noise computed with uncertainty 6.75, and
21 can then be sure that its identity is protected.

22

23 (b) Normal Additive Noise Perturbation Distributions

24

25 Now, assume a normal noise distribution, and let sd_i denote the standard deviation selected
26 by agent A_i for its noise distribution. The analysis is slightly more complicated, because
27 the $\Pr(X | A_i)$ is now a function of the position of X within the distribution, not just
28 whether it is in range as with uniform distributions.

29

30 In this case an agent chooses the standard deviation for its Normal distribution by assuming
31 that the randomized value $X' = X$, i.e. that the random noise distribution generates zero

noise. This is the worst-case, maximizing $\Pr(A_6 | X)$ because $\Pr(x | A_6)$ is maximized and the other terms in the Bayesian expansion are invariant.

10.2 Rounding and Binning of Data as an Alternative to Random Perturbation

An agent can gain the same benefits of random perturbation in a more static scheme, where data is rounded or “binned” into intervals, i.e. agents report only approximate information to other agents. This works for the same reason, mapping each released data value to more than one true value.

The problem is to select buckets for information (i.e. intervals on data) so that there are enough data records in each bucket to provide the required level of protection against adversaries aimed at uncovering the identity of agents from the information that they report.

The previous analysis extends to this case, but can be simplified. If an adversary knows that an agent has true value X , then its first task is to find the prototype value (or bucket) corresponding to the true value. The probability that an agent has the true value is the same for all agents that have posted values with the prototype value, and equal to $1/N_p$ where N_p is the number of agents with the prototype value.

The decision for agents about submitting information is also simplified: if $(1/N_p) < \epsilon$ for the prototype that the agent’s value maps into and a desired level of protection ϵ against an adversary, then select the prototype value in the interval and report information, otherwise report no information.

In order to support agents with different levels of privacy-sensitivity the system of SDI can allow agents to select a degree of accuracy when submitting information--- providing nested buckets. For example, we can allow an agent to specify that its true value can lie in one of a number of buckets, or equivalently in a single super bucket. The number of agents with true values in a particular interval is then computed as the sum of the agents that report values in an interval, and a weighted sum of the agents that state that their values are in one of a

1 This technique of binning information, or rounding information, is simpler to implement
 2 that random noise perturbation because probability distributions must neither be sent to the
 3 provider or maintained by the provider. However, the information is represented with a
 4 static accuracy which is hard to improve as the amount of data submitted increases. This
 5 occurs automatically in the standard version.

6

7 10.3 Adding Random Perturbations to Discrete Data

8

9 Adding noise to discrete values is a little more complicated, because the new value must
 10 remain feasible. For example, in randomizing the name of a CD the new name must be the
 11 name of another CD, not some “made up” name. We need to add noise to make data
 12 elements “close” to the accurate values. With discrete data, such as the name of an artist,
 13 “close” must be defined within the correct metric. The appropriate metric is such that a
 14 “close” value shares many of the same characteristics. For example, it is not appropriate to
 15 assign a close value on the basis of a shared last letter in the first name, but it is appropriate
 16 to assign a close value on the basis of an artist from the same genre of music --- from a
 17 “semantic cluster”.

18 Agents use discrete probability distributions to randomize data points.

19

20 For example, suppose that an agent wishes to reveal the name of the artist that recorded the
 21 last compact disk that a consumer purchased. An artist’s name is best viewed as a point in
 22 “artist space”, and therefore as a discrete value. It does not make sense to change a random
 23 set of letters, because the new “name” will not be the name of a valid artist. Instead, the
 24 concept of noise is to randomly choose a new name close to the current name, i.e. choose a
 25 new feasible location in artist space. One reasonable solution for the names of recording
 26 artists is to define a neighborhood of artists that are close to the original artist, and select a
 27 new artist from within the neighborhood with equal probability. The metric that defines how
 28 close artists must be computed using a system that is common knowledge to the provider of
 29 information and the consumer agent, because the provider of information must be able to
 30 compute the distribution over true artist name, given a randomized name and probability
 31 distribution. We can define the probability distribution with a threshold in the distance

1 d(name1, name2) metric, and the tag 'uniform' to indicate that each artist within the
2 neighborhood was assigned an equal probability weight. A metric for artist names can be
3 computed using clustering techniques, where artists are clustered according to the buying
4 habits of consumers that purchase them.

5
6 Semantic clustering that enables useful randomization of discrete field can be automated
7 when goods are frequent purchase, high volume goods-where individuals purchase goods on
8 multiple occasions, and more than one of the family of goods on a single occasion.
9 However, it is possible that high price, low volume goods, should be randomized on the
10 basis of expert analysis (for example new cars, computers...) -- where an expert can extract
11 key features of a purchase, and represent the purchase generically using either a single
12 prototype good, or one of a set of approximately equivalent goods.

13
14 Another approach to discrete data is to apply the "approximation" technique and allow
15 agents to specify a genre of music for example, or a prototype value from a semantic cluster.
16 This is equivalent to stating a range of salaries. For example, the system of SDI can provide
17 clusters for music, and allow a submitting agent to state the cluster of the artist for the CD
18 that it just purchased instead of the name of the artist.

19 20 10.4 Noise with Bundles of Information.

21
22 In general agents will release more than one piece of information to other agents, in a bundle
23 of information. Intuitively, when agents submit a bundle of information they open up more
24 possibilities for adversaries to determine their identity, because a bundle of independent
25 information reveals more about a user than a single piece of information.

26
27 With a couple of assumptions, we can apply the same technique of random perturbations. It
28 is again required that an agent either has information about the distribution over data points
29 present in the data base, or in the population at large. For example, consider an agent that
30 has already submitted approximate information to a provider about its home ZIP code,

10.4.1 Example

Here is a simple example for a problem where the data is two-dimensional, with two features. Consider feature a, 'year of birth', and feature b 'car model'. Suppose that agent 1 has submitted car model [VW Beetle, 0.1] to indicate that it selected a random car model from within a proximity threshold of 0.1; agent 2 has submitted car mode [Mercury Mistique, 0.05] and year of birth [1962 1968] to indicate that the year is somewhere in that range; and agent 3 that has submitted year of birth [1965 1975]. Now, agent Z, with true information (VW Beetle, 1972), computes a randomized data entry as agent 4, based on a privacy threshold of $\epsilon = 0.4$ for an adversary with correct information about agent Z. Assuming that the Mystique is not within a threshold of 0.05 of the Beetle, agent 2 has placed on probability on its true car model being the Beetle. Also, assume that the proximity threshold of 0.1 on the Beetle implies that agent 1 has a Beetle car with $\Pr(\text{Beetle} \mid \text{Agent}_1) = 0.2$. Finally, the probability that an agent over the entire population has a Beetle is 0.1. Furthermore, the probability that an agent in the population was born in 1972 is $1/(7 + 11) = 1/18$. Therefore, agent Z must place probability P_z on (Beetle, 1972) such that $P_z / (1/18 * 0.2) < 0.4$, i.e. $P_z < 1/225$. To support this, the agent can randomize each data point with probability $(1/225)^{(1/2)} = 1/15$, i.e. randomize 1972 with a uniform distribution of years $[-7, +7]$, and select a threshold on cars greater than 0.1 to give a smaller probability than the 0.2 achieved by agent 1 with a proximity of 0.1.

10.5 Bootstrapping A Data Set

Bootstrapping is an important technique for the system of strong data perturbation, in which we assume that an adversary knows that a particular agent has submitted one of the data values in a data base, and knows the true values of that agent.

A problem occurs at the start of a system, when there is little data in the database. In this case it is often possible that an agent cannot reveal any information because no other agent's data can possibly represent the agent's true value. For example, if an agent with age 18 is deciding a level of randomization, and there is currently a handful of ages in the database: {

1 [24, 28], [34, 38], [42, 46], [42, 44] } then the agent cannot submit a randomized value and
2 distribution, because the range of implied true ages will always include '18', and be the only
3 approximate information to include '18'. The agent cannot provide its age to the database.

4
5 We suggest a simple technique to bootstrap the system. The method works in the case that
6 there are a number of different reasons to decline to provide information to another agent in
7 response to its query. An agent provides randomized data and a privacy level ϵ to the
8 database, e.g. ([16, 20], 0.2), where [16, 20] is the range of possibilities for an agent's true
9 age, and 0.2 is the desired privacy level.

10
11 The SDI database can be configured to only release information when it can verify that it is
12 safe to. Similarly, an automatic method on the client machine of a user can be configured to
13 only release information when it is safe to do so. The problem that is solved here is that
14 sometimes there is no reasonable amount of noise that can be added to a data value to
15 protect a user's identity.

16
17 Another technique to solve the bootstrapping problem could hold all data until it is safe to
18 release information for the majority of data points, given the state privacy requirements of
19 users etc. The measure of "safe" is defined by the information in the database and the
20 privacy level specified by each submitting agent. A more advanced technique could
21 introduce data periodically as new parts of the data space become populated, so that there is
22 never an "active" under-populated part of the space.

23 24 10.6 Verifying Noise Levels with Playback

25
26 We can use a cryptographic technique to verify the distribution of noise that is added to data
27 - and also to enable replay. "Playback ability" - the ability to reconstruct the original record
28 from a noisy version of that record is important for a number of purposes. An individual
29 may want to obtain proof of a transaction for legal purposes and law enforcement agencies
30 with appropriate warrants might want to examine original records.

1 The client-side SDI proxy, or the central data warehouse, must keep a record of the non-
2 randomized data **X**, and a method to verify that the randomized data was computed from
3 data **X**. To generate a random noise value from a distribution **f** the agent can use a one-way
4 “trapdoor” function **trap** on the object **X** to generate a seed for a pseudo-random number
5 generator. The pseudo-random number generator then generates a sequence of random
6 numbers that are used to create the random perturbation from a well-defined algorithm.
7 Suitable trapdoor functions known in the art include the RSA encryption-decryption
8 function. It is not necessary for all agents to use the same private/public keys.

9
10 With this method it is possible to verify randomization and audit data release with stated
11 levels of randomization. SDI can request that an agent provides: **X**, **f**, and the trapdoor
12 function **trap**, and can use this information to validate **X'**. Note, that because **trap** is a one-
13 way function the agent cannot fabricate different values of **X** that would give **X'** given
14 random function **f**. SDI can “playback” the noise perturbation and produce the noisy record
15 from the original record.

17 **11. Architectural Variations**

18
19 There may be commercial contexts in which an SDI service can be established where
20 there is already in place a pre-existing trust relationship between multiple vendors and a
21 third party. Such third parties are inherently motivated to provide services to enhance
22 advertising and e-commerce for their existing and potential customers. These third
23 parties may include, for example, web hosts or e-commerce service providers (ESPS)
24 which often have hundreds or thousands of sites which they host, Web portals,
25 information and commerce service manufacturers, advertising and affiliate network
26 services and data analysis and business intelligence tool providers (which includes the
27 business to business application).

28
29 A third party may wish to implement an SDI which operates separately and
30 independently from the central SDI service. Alternatively, some of these third parties
31 may install an SDI server on their customer information server. The server may be

1 which are advertised upon with regards to what types of sites and advertisers
 2 (respectively) they allow or disallow for purposes of standard or affiliate advertising, in
 3 accordance with the methods herein disclosed. The general implementation for
 4 determining which this general application for using collective user feedback to
 5 determine relevant site links was described in the parent issued patent). In this case, end
 6 users who are subscribed to SDI would receive personalized affiliate links (including
 7 product level recommendations for on-site purchases) which have been pooled and
 8 profiled at the main SDI server from all SDI vendors (in distributed fashion) and matched
 9 with the user. Alternatively, such approach may also be used for ISPs who wish to
 10 operate their own closed user-centric SDI.

11

12 11.3 Interoperability Between Local SDI Services

13

14 With interoperable (local) SDI services, we can also facilitate the secure enforcement of
 15 data sharing policies and transfer of transaction fees between these local SDI services
 16 E.g., by/between aggregations of ad networks, syndication networks e-commerce site,
 17 portals and Web hosts operating virtual portals and advertising/syndication networks with
 18 personalization as its primary capability.

19 In each of these primary example domains, the server operator is financially motivated to
 20 sell the SDI services to his/her sites because the transaction based model is used, and the
 21 server operator receives the commission on each transaction (or click through) occurring
 22 within his/her network of sites. However, if the server operator also integrates his/her
 23 local SDI service into the main SDI service (to share user lists and impressions and/or
 24 space to advertise to these target users), s/he can receive a commission (in conjunction
 25 with each vendor transacted with) for each advertisement placement or syndicated
 26 transaction to or from his/her network.

27 We can also allow the local server operator to split the transaction fee (normally received
 28 from the main SDI service), thus "referral fee" for both the referred customer and the
 29 referral of customers (through the placement of outside ads or products on one of his/her
 30 sites) or other means of targeting his/her site's existing customers.

31

1 Figure 2 shows a couple of users connected to clients, that are in turn connected to the
2 Internet through a local intranet, such as the network of an Internet Service Provider (ISP).

3 The proxy "washes" outgoing messages of any information that would compromise a user's
4 pseudonymity, for example the "referral" field that contains the previous URL of a user in a
5 HTTP message. HTTP messages also leak other information, for example browser software
6 on a user's client machine, the operating system and a user's IP address.

7 A user can receive electronic mail through the PID and associated IP address of the ISP-
8 level proxy server.

9 The preferred implementation of this system allows the user to periodically check for new
10 mail. The client-level proxy gains access to the mail box that is associated with a
11 pseudonym by providing a correct response (signature) to an ISP-generated challenge.
12 Notice that with this solution, the ISP-level proxy has no way to connect the pseudonyms of
13 a user, so long as the user's client is not identified in its messages to the ISP-level proxy
14 server other than by the PID that the proxy makes a request for.

15 We can extend this mechanism using a technique taught in the Lucent Personalized Web
16 Assistant (LPWA). The LPWA [BGGMM 97; BGGMM98] provides for a sequential access
17 mechanism to the mailboxes that belong to a user through a one-way function that takes the
18 user's SDI log-in name and password, and an integer from 1 to N, and computes the
19 mailbox location. The mail server does not need to maintain a list of pseudonyms for each
20 user, because the user is able to efficiently access all of its mailboxes sequentially as a
21 function of other information.

1 Another variation, that relies on the user placing trust in the ISP-level proxy server, provides
2 the ISP-level proxy with the e-mail address for each pseudonym. This push method is more
3 efficient, because the ISP proxy and the client proxy communicate only when new
4 messages arrive, but provides the ISP proxy with information to compute all the
5 pseudonyms for a single user—probably undesirable.

6

7 12.2 Support for Pseudonymous Physical Mail

8

9 12.2.1 Vendor to User

10

11 A vendor must hold a “physical mail certificate” to be able to send mail (packages,
12 letters) to a user under a pseudonym. The certificate is similar to the “electronic mail
13 certificate”, in that it is signed by the private key of the user’s pseudonym, and indicates
14 that the vendor with public key P^*V can send mail to the user (under the pseudonym).

15 Each user has a trusted physical address authority, just as it has a trusted electronic mail
16 authority (the second-level proxy server), that maintains the physical mailing address for
17 each pseudonym. When a vendor has a letter X to mail to user with public key PKP , the
18 vendor generates a unique ID for the package, IDX , and sends the ID code and the
19 physical mail certificate to the trusted physical address authority of the user.

20 The physical address authority receives the certificate, $S(PKP, PK^*V, SEND_MAIL)$,
21 SKP), that indicates that the vendor is authorized to send mail to the pseudonym, and the
22 packages identify code, signed by the vendor to certify that the vendor holds the secret
23 key that matches the public key in the physical mail certificate.

24 The vendor then passes the letter X and the signed ID code to a trusted mailer, that
25 supports pseudonymous mailing, and has been certified by the central SDI server as such.
26 The trusted mailer then provides the signed ID code to the physical address authority,
27 signed with the private key of the trusted mailer. The physical address authority verifies
28 that the trusted mailer is a valid service, and releases the real address of the user to the
29 mailer. The mailer now has the letter X that the vendor wants to send to the user with
30 pseudonym P , and the physical mailing address of the user - and the package can be
31 mailed. At not time did the vendor determine the true mailing address of the user, unless
32 it works in collusion with the trusted mailer, but the trusted mailer is certified by SDI,
33 and also audited by the chosen physical address authority of the user. The address
34 authority will only release addresses to reputable pseudonymous physical mail agents.

35 We can operate physical mailing lists in the same way, and gain additional security by
36 never releasing the pseudonyms or the mailing addresses to the vendor that has requested

0969098 "102700

1 the targeted solicitations. We can use a technique that is similar to the technique that we
2 used for virtual mailing lists. The vendor describes its solicitation to the central Secure
3 Data Interchange, which leverages as much data as possible (without violating the
4 privacy policies of any of the users or vendors that are represented within the data). The
5 central SDI server generates a list of suitable pseudonyms, and then provides a series of
6 unique codes to the vendor, that the vendor can supply to its chosen pseudonymous
7 mailer with the material that is to be mailed. The central SDI server also provides the
8 appropriate address authorities with authorization to release the physical mail addresses
9 to the mailer when presented with the IDs. Notice that at no stage did the vendor have the
10 pseudonyms or the mailing addresses. The parties all have only as much information as is
11 necessary - the vendor needs some way to identify its packages to the pseudonymous
12 mailer. The mailer needs an identifier to present to the address authority, and receives the
13 addresses. The address authority just needs to know what addresses to release and to
14 which third party.

15 **12.2.2 User to Vendor mail**

16
17 The Secure Data Interchange system also provides a mechanism for users to send
18 physical mail to vendors that are registered with SDI with pseudonymous return
19 addresses. In particular, when a user sends mail to a vendor, the first-level proxy server
20 provides a tool that: (1) Computes/Looks-up the appropriate pseudonym for the user with
21 this vendor. (2) Generates a unique ID, and submits a signed message to the central SDI-
22 server, where the message relates the pseudonym, the vendor, and the ID. (3) Provides
23 the unique ID to the user.

24 The user writes the unique ID on the envelope, and mails it to the vendor. Should the
25 vendor wish to reply to the user, then the vendor can take the envelope to a
26 pseudonymous mailer, and request that the envelope be mailed appropriately. The
27 pseudonymous mailer verifies the identity of the vendor, and then submits the ID,
28 together with the vendor's signature, and its own signature, to the physical address
29 authority that is maintained by SDI. SDI releases the address to the mailer that can then
30 return the mail.

31 **12.3 Pseudonymous Payment Mechanisms**

32
33 The Secure Data Interchange architecture must be able to support all the standard electronic
34 commerce functions that we take for granted, but while maintaining pseudonymity for users
35 and following privacy policies. There are various different solutions to this problem.

36

37 **12.3.1 Anonymous Credit Card Payment [LMP 94]**

38

12.4 Client-Side SDI Proxy

The client-level SDI proxy, implemented as a client program running on the user's client machine, manages all data transfer between the client machine (and the user), and other vendors and the central SDI data warehouse. A key function of the client-level proxy is to implement profile management for a user, to control the ability of agents to track a user as he/she interacts with multiple vendors. The client-level proxy also controls release of profile information: the addition of demographic and other personal information to profiles, and the control of random perturbation to fields to prevent linking across user profiles.

The client-level proxy maintains profile information for a user's collection of pseudonyms, and allows the user to view and challenge profile information. The proxy also provides a rule-based interface to allow a user to select appropriate privacy/personalization policies.

The primary mechanism that protects the identity of a user across multiple vendors and service providers is the ability to interact pseudonymously with vendors. The user can choose a unique pseudonym for each third party with which he/she interacts, and be absolutely certain that he/she is the only party that knows his/her true identity. There is no way that a vendor can know anything about the transactions that a user has had with other vendors under alternate pseudonyms unless the user chooses to disclose the equivalence of pseudonyms, or use the same pseudonym across multiple vendors.

It is useful to distinguish three key modes of use of the Secure Data Interchange system:

- Static data-mining. Query execution with no dynamic requests for new information from information providing agents.
- Interactive data-mining. Query execution that includes dynamic attempts to request additional information from information providing agents.

- Client-side data-mining. Query execution that is performed on the local client machine of an agent, based on data stored exclusively on that machine.

All three modes may use distributed information, i.e. it is possible that the information is stored in the central SDI database, or on distributed client-side information servers, or in third-party servers. The first 'static' case and the second 'interactive' case are distinguished from the third 'client-side' case in that the query execution is performed centrally in the SDI data warehouse in the former, and on an agent's client machine itself in the latter. Client-side data mining has particular application to privacy-protected customization of information and services in on-line business-to-consumer applications.

1. Static Data Mining

In static data mining queries are executed on the information that is currently present in the SDI data warehouse, and there is no opportunity to contact agents and request more information. Applications of this type of static data mining include all types of "standard" database queries, where it is assumed that the data set is static. Queries may be open-ended, i.e. "find me all data records of this type, and perform the following operations.." or closed, i.e. "perform the following query on the data record for agent with pseudonym P1". Queries may also have side-effects, i.e. "find all records that satisfy this constraint, and then take action A".

2. Interactive Data Mining

In interactive data mining multiple agents may be contacted by the system of Secure Data Interchange in the process of executing a query, to request new information from agents, or push information to agents. A central application of this mode of data mining is *matchmaking*, which is a process where information flows between agents if and only if both the profiles of both agents are mutually compatible. In interactive data mining the querying agent does not need to know the identities of agents that are contacted by SDI, this is all transparent, and hidden from the querying agent.

3. Client-Side Data Mining

In client-side data mining the querying agent executes a query with the information associated with a specific agent, and the information remains located on that agent's client machine throughout the query. The query is executed by providing the *query method* to the client machine, processing the method with local information, and then returning a response or taking an appropriate action (e.g. displaying product X for price Y). A central application of client-side data mining is to *privacy-protected customization*, where a vendor wishes to customize its products and services for a particular agent that is registered with SDI, and take advantage of personal information relating to that agent that is not generally available. As another application, we describe a client-side advertising auction, where advertisers compete for the right to display a banner advert to a user, based on local information about the user's preferences.

4. Applications: Very Brief Overview

In the next section of the SDI description we describe some specific variations and systems that can be implemented within the general architecture. In overview, we describe the following key applications:

- Safe user profiling and personalization. This allows on-line users to receive personalized information and services without providing personal information to vendors, so that users retain control over their personal information. Users can interact with vendors under different pseudonyms, and provide information to the central SDI data warehouse to allow data mining. Finally, users can allow specific vendors to execute queries, where the result of the query is information that allows that vendor to customize its service.
- Client-side user profiling. The client-side SDI proxy can monitor the browsing behavior of a user, and submit data periodically to the central SDI data warehouse with appropriate meta information to provide the user with guarantees about the type of information that can be released to vendors during query execution.

- 1 • Static Data mining applications. Agents can submit queries to the data warehouse
2 query-execution module and perform data mining and collaborative filtering on
3 aggregated and anonymous information provided by data submitted to the
4 warehouse by agents.
- 5 • Interactive data mining applications. Agents can request actions from the SDI
6 system if certain conditions are found to exist in information, for example SDI
7 can send information to other agents about services or products, if good matches
8 are found. Another example is a request that a certain number of agents with
9 particular properties be contacted and asked to take a particular action, we give an
10 example within a transportation domain.
- 11 • Pro-active data mining. The system of SDI might itself pro-actively execute data
12 mining queries, and index and classify certain types of data to allow more
13 efficient future query execution, and to also suggest useful information to its
14 client agents. The answers to popular queries can be priced and cached, so that
15 providing agents receive value whenever an answer is sold to another agent. The
16 system of SDI might advertise a set of queries to allow vendors to select pre-
17 computed results.
- 18 • Matchmaking applications. We described in the top-level description of SDI a
19 technique to implement “persistent queries”, which reside on the central data
20 warehouse’s query execution module, and are triggered whenever the correct
21 conditions exist in the data base. One typical use of such a query is to say
22 “introduce me to other users with property P”, such that the system introduces
23 user A_1 with user A_2 if A_1 has the property required by A_2 and also A_2 has
24 the property required by A_1. This is equivalent to “introduction by mutual
25 consent”, and is possible within SDI without *information leakage*, because the
26 only agents that are informed of a match are the agents with the correct properties.
27 Applications exist to finding a business partner, funding a new start-up (incubator
28 Co.), forming an interest group, n-way negotiation, introducer system (by mutual
29 consent).
- 30

1. Static Data Mining Applications

This section describes specific applications of SDI-based static data mining. A central example is collaborative filtering and personalization applications in electronic commerce, where consumers and vendors provide information to SDI, and the information can be queried within the price and data-access rules placed with the data by owners of the information. We refer to the variation of SDI with consumer profile information as the “iamworthit” system, because users can place profile information in a shared database and receive payments in return for queries performed by vendors. Vendors can use the profile information to build better customization models, and provide customized products to customers based on their profiles and what has worked with other customers with a similar profile.

We limit our attention in this section to “static” datamining, which as defined in the mid-level SDI description allows agents to query the data, but without contacting the agents to request more information. All queries are performed on the basis of the information already submitted to the database, and the rules associated with that information. Within the set of *static* queries, we do allow a vendor to identify a specific agent within a query command, so that if a user provides an identifier to a vendor then the vendor can query the database with its collaborative filtering model and determine an appropriate action to take.

It is possible to request an *action* as the result of a query, for example “send message X to all users with profile information Y”, so long as this message is not contingent on as yet unknown information about the user. We describe specific examples, for example to a smartbrowsing system, an education portal, and an advertising network.

One of the core purposes of SDI is to provide a common location and format for information that has been gathered from a wide variety of sources and that might require different sorts of analysis. Since its framework is designed to handle different types of data and algorithms, SDI can be used as a platform to explore and exploit the rich

1 share only a portion of his database, or that he will release only randomized data in
2 accordance with the level of privacy he has guaranteed his customers. Although such
3 restrictions could impact the content of the data analyzed by a vendor, as long as it is kept in
4 an SDI-compliant format it can be analyzed by SDI's suite of tools.

5 The data that is stored in the central SDI server has tight usage restrictions. For example, the
6 user will have specified (via a price-policy) that data can only be used by querying agents
7 that are able to present certain types of certificates, only anonymously, etc. The central data
8 warehouse maximizes the value of data by analyzing conditions and supporting queries with
9 the cheapest possible price rules.

10 The information that is stored in the central SDI data warehouse can be provided by users,
11 via their SDI client-level proxy agents, and relating to their own individual purchasing and
12 buying habits. Alternatively, the data might be provided by vendors, and relate to what
13 information that vendor decides to release about its customer base—and in agreement with
14 stated practices that the vendor follows about the information that it collects on its
15 customers.

16 It is not necessary for data records to physically reside on the central SDI data warehouse. A
17 vendor may provide links to data, and use a locally secured query-execution module to
18 carefully control queries of the data. For our current purposes we assume a uniform data
19 access model, where although both data and algorithms might reside either at vendors' home
20 locations or within the SDI system itself, the general analysis works transparently across
21 these boundaries.

22 For security reasons, the contents of databases may be injected with a small amount of
23 noise. This prevents database users from surreptitiously connecting database records to
24 individual customers, yet maintains the quality of inferences made about the database in
25 general. This is described earlier, as noise-perturbed data release. Although such "noisy"
26 records don't pose too much of a problem for those methods that make generalized
27 inferences, it should be noted that recommendations made for individual customer vectors
28 that have undergone such randomization will be less useful, since predictions are being
29 made for a noisy target.

30 A final consideration is the reduction of the data vectors' dimensionality (which can be
31 extremely high), since it is harder to make clean inferences about sparse data. There are

many standard methods that can be used to achieve this, such as Principal Components Analysis. Another approach is to adjust the granularity of the data, if at all possible. In a music store analysis, for example, there might be many more album titles than artists (since each artist can produce multiple albums). In such a case, purchases could be recorded by artist rather than by album, greatly reducing the dimension of the customer vectors' purchase space.

1.1.1 Data Structure

In this application there are many types of information which can characterize both users and items. SDI is intended to function as the intermediary between a vast web of vendors, on the one hand, and individual consumers, on the other hand. Major sources of data include:

- 1) Demographic. Such data will most likely be elicited by SDI from vendors and consumers when they initially register for the service, and details very general characteristics about them. It will consist of numbers and categorical values (age, zip code, sex, level of education, etc.).
- 2) Commercial. This is the kind of data any that vendor collects in the course of doing business (especially e-commerce); generally, it links customer codes to purchase items, dates, quantities, and prices. Depending on the nature of the business, this data could be fairly complex, and might well include text. For example, one could imagine that a bookstore, in addition to keeping track of its sales history, collects book reviews, author profiles, and plot summaries.
- 3) Behavioral (vis-à-vis the Internet). A user's client-side SDI proxy can monitor his/her browsing behavior on the World Wide Web, monitoring the pages a user hits, the click stream and content requested, etc. Click stream information can be useful, for example, because it can indicate a user's interest in the information that it is presented with.

We assume that vendors and/or a third-party annotate web pages with *tags*, that provide a commentary of a web page and allow meaning to derived from a user's browsing behavior.

004207" 8606960

1 The mere presence of such tags allows for correlations to be drawn between different web
2 pages (e.g., a common Extensible Meta Language (XML) tag used by travel-related sites),
3 because it implies similarity. Furthermore, it is conceivable that such tags could encode
4 more refined measures of a web page's content, such as browsers' evaluations of its value.
5 For example, a web page of interest to scale modelers, in addition to having images and text
6 related to model trains, might have an XML tag that shows that other scale modelers have
7 given the web site a "five-star" rating. This page should therefore be given a greater weight
8 when SDI is used to create correlations of interest to model hobbyists.

9 In what follows, we assume that SDI contains data of the following types:

- 10 1) Numerical (e.g. an age, price, or period of time).
- 11 2) Categorical (e.g. a color or musical genre).
- 12 3) Text.

13
14 We describe a mathematical model for computing useful collaborative filtering queries
15 across heterogeneous data records, such as the data stored within the central SDI data
16 warehouse. Throughout, we present only the basic mathematical model. In implementation
17 the query would be formulated in a general-purpose language, for example Java, and
18 executed in the central SDI data warehouse subject to the conditions on access to data as
19 expressing users *price-access* rules.

20 **1.1.4 Computing Correlations Between Data Points**

21
22 A common task for SDI is to compare and correlate different customers, which might well
23 be represented by mixed collections of numbers, categories, and blocks of text. This is
24 handled by treating each customer c_i as a vector in a space whose coordinates correspond to
25 the fields of data available. In the following description we refer to a customer, but when a
26 user interacts with a vendor under a pseudonym, the profile information will only relate to
27 information provided to the central SDI server for that pseudonym.

28 If there are m numerical pieces of data available, there will be n corresponding coordinates
29 in the data space, (x_1, \dots, x_m) .

30 For each category i , there will be a corresponding number of values, n_i . Hence, for a color
31 category {red, white, blue}, $n_{\text{color}} = 3$. Since each value is assigned its own coordinate,

1 category i is represented as an n_i dimensional vector, y_i . Hence, the total number of
2 dimensions used to describe the full set of n categories (y_1, \dots, y_n) is

$$\sum_{i=1}^n n_i$$

3

4 Note that sparse methods are especially useful here, since a categorical vector y_i will
5 typically consist of mostly zeroes, with a single non-zero coordinate representing the
6 categories' value (i.e., we encode the color red, using the previous example, as (1,0,0)).

7 Note also that category vectors with different values are treated as orthogonal by the system.

8 A final issue is the representation of text. As described in previous related patents, all
9 relevant blocks of text in the database are converted into a dictionary that maps unique
10 strings to the number of times they appear in the database. An appropriate TF/IDF weighting
11 function is chosen and calculated for each of the p words that appear in the dictionary. The
12 full set of text connected to a single customer can thus be represented as the vector ($z_1, \dots,$
13 z_p), where each z_i equals the number of times the word i appears in text related to the
14 particular customer multiplied by the TF/IDF score assigned to word i .

15 In summary, when a database describes its customers using a combination of numerical
16 values, categories, and text, customer i can be represented by the vector $c_i = (x_1, \dots, x_n, y_1, \dots,$
17 $y_m, z_1, \dots, z_p)$.

18 **1.1.5 An Example Profile Vector**

19

20 Suppose we have a database containing information on customers' ages, their musical
21 preferences (i.e. an answer to a survey asking: "Which do you prefer, Mozart or the
22 Beatles?"), and the contents of the emails they've written. Furthermore, suppose the only
23 salient variables in all the emails written consist of the words "Beatles", "Mozart", and
24 "practice", and that we are using the function

$$TF / IDF(x) = \frac{1}{\sqrt{n_x}}$$

25

26 Where n_x represents the number of times word x appears in the dictionary. We now want to
27 represent one of the customers in the database; he's a 10-year-old boy who prefers Mozart to
28 the Beatles, and who wrote an email to his friend that mostly describes his attempts at

1 Since the level of granularity will have a salient effect on the outcome of an analysis, it
2 should be chosen very carefully, and might well play a factor in pricing when a vendor
3 chooses to sell its data.

4 **1.1.7 Statistical Methods for Data Analysis**

5
6 In order to perform a wide range of analytical tasks, SDI needs to make use of a variety of
7 computational approaches. These are described below, starting with the simplest methods
8 first.

9 • (1). Standard Database Searches

10
11 Since most of the data will be stored in centralized databases, simple searches,
12 queries, and data filters can be implemented by means of standard SQL commands.
13 Typically, data will be collected or sorted using efficient database calls before being
14 fed through analysis routines; once complete, the results can be fed back out to the
15 database environment for further efficient manipulation.

16 • (2) Metrics – Measuring the Similarity Between Profile Vectors

17
18 Given two customer (or vendor) profiles, c_i and c_j , it is frequently desirable to know
19 how similar they are. For this purpose, we define the similarity metric $M(c_i, c_j)$ to be a
20 function that takes as input two customer vectors and returns as output a numerical
21 value in the range $[0,1]$. When two customers c_i and c_j are identical, $M(c_i, c_j)=1$; when
22 they're completely different, $M(c_i, c_j)=0$.

23 The problem is somewhat simplified by the fact that we treat all customers as vectors.

$$M(A, B) = \cos \theta = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

24 Given two customer vectors, we can use the correlation between them to serve as our
25 metric:

26 Note that θ here represents the angle between the vectors A and B , and that we expect
27 all coordinates of the vectors to be positive (in order for $M(A,B)$ to keep its output in
28 the range $[0,1]$).

In more complicated cases, however, a customer vector might contain multiple fields with varying ranges of values. For example, we might have customer vectors of the form $c_i = (\text{age}_i, \text{income}_i)$, in which the maximum age is 80, but the maximum income is 300,000. In such cases, the coordinates with larger values will dominate the similarity metric, overwhelming any influence that smaller fields might have.

This requires a normalization of the customer vectors, which can be done in several different ways. One approach would be to scale every coordinate by the maximum observed value, forcing all coordinates to lie between 0 and 1 (again, enforcing the rule that all coordinates must be positive).

$$c_i = \left(\frac{\text{age}_i}{\max(\text{age})}, \frac{\text{income}_i}{\max(\text{income})} \right)$$

The only problem with this is that if a coordinate's maximum value is an outlier (being vastly bigger than the typical value), most of the coordinates' values will seem unusually small once they are scaled by the maximum. In such cases, it might be better to scale the values with a "squashing" function such as the sigmoid, which deadens the impact of extreme values; one such configuration would be the following:

$$\begin{aligned} \overline{\text{age}_i} &= \frac{\text{age}_i - \text{mean}(\text{age})}{\sigma_{\text{age}}} \\ \overline{\text{income}_i} &= \frac{\text{income}_i - \text{mean}(\text{income})}{\sigma_{\text{income}}} \\ c_i &= \left(\frac{e^{\overline{\text{age}_i}}}{1 + e^{\overline{\text{age}_i}}}, \frac{e^{\overline{\text{income}_i}}}{1 + e^{\overline{\text{income}_i}}} \right) \end{aligned}$$

Note that the mean and variance of the data points are used to fully normalize them, such that the sigmoid function will spread the values somewhat more evenly between zero and one.

The previous approaches are especially useful for single numerical fields, which might well overwhelm each other if some sort of normalization isn't performed.

A different problem arises for text or large categorical fields, since they can potentially consist of hundreds of coordinates capable of overwhelming the influence of single numerical fields. Suppose we believe the age of a customer is as important as the text of articles read. In such a situation, the thousands of coordinates devoted to the text field would dominate the metric's behavior, negating any influence that age would have on our measure of similarity – clearly not a good situation. A solution to this would be to find the correlations among the fields taken separately, then average the result. That is, if each customer $c_i = (age_i, text_i)$, where $text_i$ is a vector with a very high number of dimensions, we could define the metric:

$$M(c_i, c_j) = \left(\frac{corr(age_i, age_j) + corr(text_i, text_j)}{2} \right)$$

Where

$$corr(c_i, c_j) = \frac{c_i \cdot c_j}{\|c_i\| \cdot \|c_j\|}$$

The result is a metric that gives equal influence to each field.

• (3) Forming Vectors Into Groups

The process of classification is essential to collaborative filtering, as it allows different vectors to be formed into groups based on some measure of similarity. If we are able to create groups of customer vectors, for example, we can then give individual customers recommendations based on the patterns of their group-mates, who presumably have similar tastes.

K-means Clustering and Nearest Neighbor algorithms are extremely useful for grouping purposes: previous iReactor patents give a full and detailed description of our customized versions. This section gives a brief overview of these methods.

(3.1) Clustering

K-means Clustering is an algorithm used to partition a coordinate space such that all vectors in a given partition are more similar to that partition's vector average (the

centroid), than to the centroids of any other partition. It is a process that iterates over the following steps:

0. "Seed" the coordinate space with the initial centroids, which are vectors used to describe the centers of the clusters, in the sense that they are the average of all the vectors currently assigned to the partition. This can be done randomly (assigning centroids random coordinates) if no other information is available, or it can be guided by pre-existing information. For example, if we wish to cluster vectors of music customers, we can use information about musical genres to create initial partitions that correspond to pop, gospel, classical, etc. This will locate the centroids in well-spaced intervals across the coordinate space.

1. Assign vectors to the most similar centroids. This is done for each vector by scanning across all centroids and calculating similarity $M(\text{vector}, \text{centroid}_i)$; once finished, the vector is assigned to the cluster whose centroid has the greatest similarity. In this stage, vectors may switch their allegiance from one centroid to another, if the relative distances to the vector have changed sufficiently since the previous iteration. If no vectors change their allegiance, the iteration process is complete, and the algorithm stops.

2. If the iteration is not complete, recalculate the centroids by setting them equal to the average of those vectors that have been assigned to them. Go back to step 1.

Once the algorithm converges, the vectors are grouped into clusters. The centroids' coordinates as well as the identity of cluster members is useful information that can be passed on to subsequent stages of analysis.

(3.2) Nearest Neighbor

The nearest neighbor algorithm, simply stated, creates a list of those vectors in a database that most resemble a particular target vector. This is accomplished by comparing the target vector, in turn, to every other vector in the database; the similarity between them is recorded, and once the comparison loop is complete the list of similarities is sorted. The top k members of this list are returned as representing those k vectors which most resemble the target.

• (4) Generalizing Across Databases

1

2 One of the most useful aspects of SDI is that it allows for inferences to be drawn
3 across different databases through underlying connections in membership or content.

4 An especially strong link can be made between commercial databases if they have
5 customers in common. However, for reasons of privacy, individual customers may
6 choose to use different pseudonyms when dealing with different vendors. This might
7 be preferred by the individuals, but it weakens the inferences that can be made
8 between fields occurring in different databases.

9 The techniques chosen to infer correlations across different databases will depend on
10 how many pseudonyms are shared in common. At one end of the spectrum, every
11 customer uses a single pseudonym for all transactions, and makes an appearance in
12 every database. At the opposite end of the spectrum, every customer uses a different
13 pseudonym with every vendor, and may appear in only a single database.

14 Case 1: All customers use a single pseudonym, and appear in all databases
15 considered.

16 This is the simplest situation to handle. Since all customers appear in all the
17 databases, the customer vectors' fields are essentially scattered across several
18 locations, but can be easily reconstructed. For each customer, we define a new data
19 vector that concatenates that customer's representation from across the different
20 databases.

21 Hence, if we are considering databases A, B, ..., Z, and customer i appears in each
22 one, we define a new vector $c_i = (c_{Ai}, c_{Bi}, \dots, c_{Zi})$, where c_{Ai} is customer i's vector in
23 database A. We then proceed as usual, making inferences with these augmented
24 customer vectors.

25 Case 2: Most customers use a unique pseudonym, and frequently appear in different
26 databases.

27 In this situation, although we see some connections between the databases, many
28 pseudonyms appear in only a single location. Using Bayesian techniques, however,
29 we can still make predictions for customer vectors across databases.

1 Suppose we have a set of databases, A, B, ..., Z. Taking each database in turn, we
2 cluster it using all available data. Thus, using every record in database A, we group
3 A's customers into clusters
4 A_1, A_2, \dots, A_n . Taking database B, we create clusters using all of B's information,
5 creating customer clusters B_1, B_2, \dots, B_m , and so forth.
6 Now, scan both databases for common pseudonyms (representing those customers
7 who have interacted with both vendors under the same pseudonym) and create count
8 variables w_{ij} to represent the number of pseudonyms that appear jointly in A_i and B_j .
9 We can now produce the probability that a pseudonym appearing in A_i will appear in
10 B_j :

$$P(B_j | A_i) = \frac{P(B_j \wedge A_i)}{P(A_i)} = \frac{w_{ij} / total}{\sum_{j=1}^m w_{ij} / total}$$

$$total = \sum_{i=1}^n \sum_{j=1}^m w_{ij}$$

11 For example, if we have a database of airline ticket purchases and a database of
12 restaurant visits, we can create clusters, in the first case, of customers who travel to
13 similar destinations, and in the second case, of customers who eat at similar
14 restaurants. Given that a particular customer belongs to a cluster of people who
15 frequent Caribbean restaurants, we can infer which travel packages would most
16 appeal to him based on the linking probabilities, as defined above.

17 • Multivariate Extensions:

18
19 If we have a third database C, and there are a large number of pseudonyms common
20 to A, B, C, the above probabilities can easily be extended. For example, knowing that
21 a customer appears in A_i and B_j , we can calculate the linking probabilities to any C_k :

22

$$P(C_k | A_i \wedge B_j) = \frac{P(A_i \wedge B_j \wedge C_k)}{P(A_i \wedge B_j)} = \frac{w_{ijk} / total}{\sum_{k=1}^p w_{ijk} / total}$$

$$total = \sum_{k=1}^p \sum_{i=1}^n \sum_{j=1}^m w_{ij}$$

1

2 Or, if there aren't many pseudonyms that span all three databases, the probability of

3 C_k given that a pseudonym exists in A_i and B_j could be approximated by:

4

$$P(C_k | A_i \wedge B_j) = P(C_k | A_i) \cdot P(C_k | B_j)$$

5

6 Case 3: All customers use several pseudonyms, and none appear in different

7 databases

8 In this situation, there are no common customer codes that can be used to create links

9 across the databases. However, the mere fact that several databases have been

10 brought together for analysis should imply that there are semantic commonalities in

11 the data.

12 Although each database contains different fields, it may be the case that those fields

13 deal with related subjects. A human expert, knowledgeable in the content of the

14 databases, the subtleties of the domain, and the overall goal of the analysis (e.g. the

15 creation of recommendations), will be in a position to create a "common-information

16 profile" that spans the databases. In essence, the common-information profile defines

17 a format that allows vectors from different databases to share a common coordinate

18 space.

19 The idea is this: the expert designs a high-level vector format that embodies the

20 content deemed important for the project goals. Next, for each database he develops a

21 mapping that encodes the database's elements into the generic format. Finally, the

22 desired analysis is performed on the full set of common-information profiles.

23 Although the expert will have to create completely new fields for the common-

24 information profile, certain types of data will map directly to the common-

25 information format. In particular, if every database contains text (catalogued and

counted, for TF/IDF purposes, by accompanying dictionaries), the union of the words will define the text coordinates of the new common-information profile. When word counts are being mapped from their original databases to the new vector, the original TF/IDF weightings may be used, or new TF/IDF weightings may be created (using a dictionary constructed from all the databases' text taken together).

Once analysis has been performed, certain common-information profiles will be grouped together by their shared similarities, although the pseudonyms they represent may have been originally drawn from different databases. Such groups will represent links between different databases, and may be used for predictive purposes (see end of example).

1.1.8 Example of Cross-database Analysis

In this example, suppose that the central SDI data warehouse contains data submitted by the following vendors:

- A. A travel agency keeps track of tickets sold, and vacation web pages browsed.
- B. A bookstore keeps track of books sold, and stores an electronic version of the New York Times Review of Books.
- C. A sporting-goods and clothing shop, keeps track of purchase items sold (which includes magazines, for which electronic text exists).

A certain airline wants to promote various vacation packages it has available, which include both European and Caribbean vacations, as well as singles and family packages. Although it has leased rights to databases A,B, and C, it turns out that no customer pseudonyms appear in more than one database at a time – in other words, there are no shared records.

A vacation expert is hired to create a common-information profile. He creates the following information vector: (list of tropical countries, list of European countries, family score, list of sports, text)

Note that the family score is a numerical value ranging from 1 (young singles) to 10 (many small children), and indicates what kind of person the customer is (a party-oriented student vs. a sedate father of three).

The expert the creates the following mappings:

- 1 A. Travel Agency. Link destinations of tickets sold to country fields (i.e., the number
- 2 of trips to Germany by a customer would be placed in the Germany field of the
- 3 common-information profile). Link sales of children's tickets, or requests for
- 4 children's meals, to family score. Put web-page data into text field.
- 5 B. Bookstore. Link travel books' text to country lists. For all books purchased by a
- 6 customer, map text from book reviews into text field.
- 7 C. Sporting-Goods store. Map warm-weather clothing (and swim gear) to tropical
- 8 countries, ski gear to countries with skiing areas. Map sales of toys or children's
- 9 clothing to high-value family scores, map revealing-bikini and student-discount
- 10 sales to low-value family scores. Map text from magazines purchased by a
- 11 customer to text field.
- 12

13 These mappings are then applied to each database, generating a full set of common
14 information profiles. These are then clustered, forming groups that share commonalities.
15 The expert can now do several things with the results. First of all, he identifies the general
16 "flavor" of each cluster (e.g., families with small children that enjoy winter, Europe, and
17 skiing); the pseudonyms contained within each cluster can then be targeted for vacation
18 packages suitable to their tastes. Secondly, the fact that pseudonyms from different
19 databases have been clustered together allows the expert to plan cross-category marketing. If
20 certain travel-book-buying parents have been grouped together with parents who bought
21 their children swimsuits and scuba toys, it may be that they share a preference for family
22 activities that take place in warm places or by the seashore. Hence, the book-users might be
23 advertised various ocean-related sports goods appropriate for young families, and likewise
24 the swimsuit-users might enjoy getting recommendations for travel books that describe
25 tropical destinations that are especially fun for children. That is, if the goal is to cross-
26 market items from A to customers in C, the most logical source of recommendations would
27 be the people in A who have been grouped with the people in C.

28 **1.1.9 Methods for Validation**

29

30 To a large degree, the overall success of an SDI analysis is the relevance of the connections
31 that are inferred from the data. It is often the case that a certain amount of validation is

adding it to a running total. The system with the highest total can thus be judged the most effective, since it most strongly recommended items that the customers did, in fact, end up purchasing.

Because the result of this type of validation is a quantitative score, it is possible to automate the model selection process. Given a set of analytical approaches (each with its own array of parameter settings), it is possible to loop through the full parameter space (using a grid of evenly spaced numerical values, if needed, to reduce dimensionality), computing a validation score at each iteration. Those combinations of algorithms and parameter settings that demonstrate the best performance could be chosen as the top candidates for the final system configuration, since they do the best job at predicting customer behaviors.

(1b) Quantitative Approaches – Dynamic Method

The problem with the hold-out approach to validation is that it isn't dynamic, since it doesn't reflect the impact that the recommendation system has on the customers once it is implemented, and may be based on data that doesn't contain current trends. After all, it is better to predict what the customer will buy rather than what the customer has bought in the past.

A better approach is to run a controlled experiment against the actual customer base. First, the pool of customers is split at random into different segments. Next, each approach under consideration is used exclusively to make predictions for a given segment. Once the trial period is over, each system is given a score based on how valuable its recommendations turned out to be (this could be measured by total sales generated, for example, or by the number of times a customer made use of a recommendation).

(2) Human Expert in the Loop

Although quantitative methods can automate the validation process to some degree, at the beginning of many projects there is so much raw input data available and so many decisions that have to be made about the analytical approach that an automated process would have to test a prohibitive number of combinations of data, algorithms, and parameter settings to get optimal results. In such cases, it is useful to employ a

1 human expert who understands the psychology and nature of the particular domain
2 being analyzed.

3 Such a person will have intuition about what is and isn't relevant for his domain. For
4 example, a movie expert might be called in to work on a movie-recommendation
5 system, for which an immense amount of input data is available. In choosing relevant
6 fields for analysis, the expert's understanding of cinema would lead him to include
7 the director's name and numbers of Oscars awarded, whereas the exact length (in
8 minutes) of the movie would be, in his estimation, irrelevant and therefore excluded.
9 Once the analysis is complete and recommendations have been made, the expert's
10 opinion (based on a qualitative understanding of the domain) can be used to guide
11 which particular combination of settings, chosen from a list of candidates with
12 detailed test outputs, should be used for the recommendation system.

13 (3) Combined Human/Quantitative Method

14
15 There is certainly no reason why both approaches couldn't be used in combination.
16 Many data sets include fields that are extremely noisy or simply irrelevant to a given
17 problem; a human expert can be employed to pare the data set down to a reasonable
18 size and dimensionality, using his domain expertise to create a data model reasonable
19 for the proposed analysis. Next, automated methods can be used to fine-tune the
20 parameter settings and to choose which subsets of the input data are the most useful.
21 Finally, the human analysts called back to qualitatively evaluate the results of the
22 fine-tuning, making the decision to either start a new iteration of the analysis, or to
23 certify that the process is complete and ready for commercial application.

24 1.2 The Iamworthit System

25
26 Iamworthit is an application of Secure Data Interchange to business-to-consumer (B2C) e-
27 commerce, where individuals perform transactions on-line through interactions with the
28 server machines of on-line vendors. Individuals interface with the servers of vendors via
29 local user client machines, networked at present over the Internet, although this is not a
30 necessary infrastructure for the invention.

1 Iamworthit allows individuals to receive payments for the information that is collected by an
2 SDI client proxy and stored in the central SDI data warehouse, for data mining purposes.
3 Iamworthit also allows individuals to receive payments in one-to-one interactions with
4 vendors in return for providing vendors with information that allows them to make an
5 appropriate offer. Secure data interchange supports the useful exchange of information
6 between agents without allowing vendors to collect and distribute information about users
7 without the knowledge or permission of users, as is possible now via the system of cookies
8 and affiliate networks, such as that operated by www.doubleclick.com. In the system of
9 Iamworthit vendors benefit through well-targeted advertising (both push and pull), the
10 ability to customize information and services (even to first-time customers), and access to a
11 large database of information about buyer purchasing habits. Individuals still release profile
12 information to the central SDI database for the purposes of controlled access by vendors,
13 and also provide vendors with profile information during interactions and allow vendors to
14 execute queries on information and receive the benefits of personalization without directly
15 accessing the information. Users can maintain multiple pseudonyms and profiles, but within
16 SDI vendors can still access information across pseudonyms and use cross-web (broad) and
17 single-vendor (deep) information to build robust models of buyer behavior. Buyers benefit
18 through personalization with privacy, and financial rewards in return for releasing profile
19 information.

20 Within B2C e-commerce, secure data interchange can also support a system of *time-of-*
21 *purchase competition*, which allows an individual to use the profile management capability
22 of SDI to provide vendors with information about a user at the time of purchase, and allow
23 competing vendors to offer the same product at a better price, or a better product (for the
24 user) at a good price. With time of purchase small entrants to the marketplace can make
25 counteroffers to users that are about to purchase a product or service from another vendor,
26 and can compete in small parts of the marketplace without investing heavily in advertising
27 and brand awareness. Time-of-purchase requests can be made by user clients to iamworthit,
28 to request that iamworthit cascades purchase requests onto other vendors, collects responses,
29 and then return them to user clients. This extension of SDI is discussed in Section 2.1.

30 As a commercial strategy, one might make it a necessary condition of belonging to
31 Iamworthit that client machines submit profile information to the SDI central data

warehouse, so that vendors can perform useful data mining and then provide customized products and information to users, for example based on models of collaborative filtering.

1.3.1 System Overview

Client machines in Iamworthit implement an SDI proxy on-top of a user's regular web browser. The role of the proxy is:

- a) Profile and identity management as the individual interacts with server computers of on-line vendors.
- b) Collect and manage profile information, with information periodically submitted to the central SDI data warehouse.
- c) Release profile information to vendors according to profile management policies.

In Iamworthit the client-side SDI proxy is configured by users to periodically push information collected about the user to the central SDI data server, i.e. click stream data, profile information, purchases made, information requested, etc. Profile information is associated with price-rules, as in the top-level description of SDI, and can also be adjusted with random perturbations to protect the identity of a user under multiple pseudonyms.

The client periodically sends update messages to the database, for example with information about new activity (e.g. web browsing, purchases, adverts not selected, etc.) SDI allows the information to be used by vendors to build models that enable good personalization of products, services and adverts.

Users might choose to randomize profile information to protect sensitive information and to prevent vendors linking users across different pseudonyms. Clients submit perturbed data points and the distribution that was used to generate the noise (see the section on random noise perturbation).

1.3.2 Privacy-protected data mining

Vendors can gain access to profile information in the central SDI data warehouse, as long as they meet the criteria of the agents that submit information on behalf of users, and pay the price of data access. The profile information can allow vendors to build complex models for personalization, negotiation, and advertising, based on information about purchases made by different types of users.

1 Users submit profile information to a central database that can be used for profiling, without
2 revealing their identities. Client-side SDI proxies track user activity on-line, across multiple
3 sites, and submits data, randomized if necessary, and only according to a user's preferences
4 to the central SDI data server. The data is useful for building models of buyers, for example
5 purchasing patterns, for the purposes of personalization of information and adverts. Agents
6 that submit information retain ownership of the data.

7 Simple data mining queries include:

8 (a) Compute the average income level of people purchasing camcorders. Suppose
9 that a vendor has pseudonymous identities of its recent customers, and wants to
10 compute their average salary. It does not know the salary of any of the customers,
11 and individual customers will not release their salary to the vendor. However, if
12 some of those customers have provided information about their salary to the
13 central SDI data warehouse then the vendor can compute the result to its query.
14 Agents might associate price rules with information about their salary that allow a
15 vendor to query that information so long as the identity of the agent is not
16 revealed, i.e. so long as the salary is provided anonymously. We described earlier
17 in the patent description how price rules can define different prices for different
18 types of information access.

19
20 (b) Compute the total donation to a fund for bone cancer research. Individuals do not
21 wish to release information about their specific charitable donations, but might be
22 happy to release that information anonymously. Again, a query to compute an
23 average donation can be formulated and executed in the central SDI data
24 warehouse.

25
26
27 (c) Perform collaborative filtering across multiple fields. This is explained in more
28 detail below, and is possible within SDI because a general query can be
29 performed so long as the querying agent makes payments for data access in
30 accordance with the price rules of agents which submit information to the shared
31 database. Note also that randomly perturbed data does not prevent the

1 computation of correlations between fields, so long as the random perturbation is
2 “small” with respect to the value of the data record.

3 **1.3.3 Importing Off-line Data into the Data warehouse**

4
5 The central SDI server can associate off-line information about a user with a user’s on line
6 pseudonymous profile, even though the central server does not know the user’s pseudonym
7 IDs. This can only be done with the user’s consent, and may also involve appropriate
8 compensation. Within the system of iamworthit we can credit users for both off-line and on-
9 line information.

10 Merging a marketing database with SDI user profiles can be useful both to initialize the
11 database, for example when asking a user questions to generate an accurate user profile
12 rapidly and efficiently. Off-line data can also add useful richness to on-line profiling
13 information, which may be largely contextual and low on details/factual information. For
14 example, off-line data can include information such as whether a user owns a car, rents an
15 apartment, has house insurance, life insurance etc. SDI can also extrapolate correlations to
16 other user profiles, on the basis of common SDI-profiles, for example using statistical
17 techniques.

18 It is often the case that individual customers appear in some databases, but not in others.
19 Under normal circumstances, an analyst working across different databases would be faced
20 with a large number of incomplete customer records, each with gaps corresponding to the
21 fields of the databases to which they don’t belong. A solution to this problem is offered by
22 SDI, which is capable of drawing correlations between different databases - this information
23 can be used to generate predictions to fill in the gaps of incomplete customer records. The
24 result is a full set of customer records that can be meaningfully sorted or filtered by any of
25 the combined fields, and which can now be handled as a unified set of data, suitable for use
26 by standard database analysis systems.

27 In a typical example, SDI might be used to combine a demographic database, such as the
28 one offered by the Econometrics Corporation, with a commercial database, such as the one
29 offered by Claritas. The Econometrics database consists of 180 million different customer
30 records, but at a fairly coarse-grained level of detail, consisting of such information as age,
31 gender, family status, location (at the state, city, or zip code level), and personal income. In

1 comparison, Claritas offers a smaller base of customers, but includes information of
2 arguably higher quality, since it breaks customers down to the geocode (sub-neighborhood)
3 level, and includes much more detailed information on personal spending habits across
4 hundreds of different purchase categories. A logical reason to combine these databases
5 would be to supplement information about customers in the vastly broader demographics
6 dataset with particular predictions about their personal preferences and likely commercial
7 spending habits. One could imagine using this augmented data set to support a web site that
8 instantly customizes itself to new visitors' preferences. Since the number of records in the
9 Econometrics database is equivalent to roughly 72% of the population of the United States,
10 it is likely that most first-time visitors to the site will already have a "thumbnail sketch" in
11 the system, and can thus be greeted with an page appropriately configured to their personal
12 tastes.

13 The technical details of the combination process (which have been described elsewhere in
14 the patent) to a large degree depend on the amount of overlap between the databases, that is,
15 the number of customer records which are shared in common.

16 Suppose the demographic databases' fields are coded (x_1, \dots, x_n) , and the commercial
17 databases' fields are coded (y_1, \dots, y_n) . Suppose further that customers in set A appear only in
18 the demographic database, customers in set B appear only in the commercial database, and
19 customers in set C appear in both.

20 The process of supplementing the fields of customers A depends completely on the
21 derivation of the distribution $f(y_1, \dots, y_n \mid x_1, \dots, x_n)$, which describes the correlation of fields
22 in the commercial database on fields in the demographic database. As previously discussed
23 in the patent, different techniques may be used to create this distribution, depending on the
24 size and variety of C.

25 As a concrete example, one could imagine that set C includes customers from rural areas.
26 The demographic database would reveal that, although their incomes aren't huge relative to
27 the national average, they tend to spend a lot of it (i.e. are active consumers), have large
28 families, and purchase large vehicles. The commercial database might show that they
29 enjoying hunting magazines and Ford trucks. If they live inland, they buy hunting
30 equipment, if they live near the ocean, fishing equipment.

1 If these trends are dominant in set C, they will impact the distribution function. Thus, when
2 a browser from a small town in Texas with a typical income pattern visits the automated
3 website, he could be greeted with discounts on truck accessories and a small sidebar with
4 news on the hunting season. On the other hand, a visitor from a small town in Maine might
5 be given the same truck discounts, but would have news on the fishing season.

6 Although the demographic dataset is arguably the weaker of the two in terms of content, the
7 fact that it contains even a small amount of information on most people in America makes it
8 very valuable for handling first-time visitors, since most of them will appear in it. By using
9 SDI to leverage the more detailed information in the commercial database, we are able to
10 supplement the rough demographic data with predicted commercial preferences. This allows
11 us to construct more detailed thumbnail sketches for each customer, allowing our reception
12 of first-time visitors to be much more appropriate (since knowing personal hobbies or
13 interests tells us much more about a person than general income level).

14 1.3 Static Query Execution: Central Applications

15

16 Some key applications of SDI are:

17

- 18 • Assessing the Value of Data. Plug together sets of data, and measure predictive
19 accuracy.
- 20
- 21 • Matching Data Across Vendors. Find patterns in common pseudonyms, denoting
22 common areas of interest; use catalogues of order codes and item description to
23 find similarities across data sets.
- 24
- 25 • Targeted Recommendations; e.g. match customers to their nearest neighbors in a
26 data set and generate recommendations for users, collaborative-filtering style
27 application.
- 28
- 29 • Leveraging Portal Data. Use data from portal to leverage data needs for ISP

30

- 004207-886666
- 1 • Analyzing Affinities. Suppose a vendor has a list of customers, and knows to
2 some degree what web pages they visited after leaving vendor site. A large
3 collection of customers taken from an ISP will contain their web-surfing behavior.
4 Cluster web sites and cluster customers, finding cluster-to-cluster interactions.
5 Use this information to classify vendor's customers; gives vendor an edge in
6 knowing customers' tastes.

7
8 There are a number of search-based applications, where SDI searches for appropriate
9 profiles and then requests that SDI makes contact with the users pseudonymously, i.e.
10 without the vendor receiving any useful information about a user's identity. The contact,
11 interaction, and business relationship with the vendor occurs under terms of complete buyer
12 pseudonymity. In accordance with the parent patent application [INSERT US PATENT
13 No.] the pseudonymous communication may be either email, real-time text
14 communications, voice (such as the pseudonymous telephony or Internet telephony). In the
15 case of pseudonymous telephony, instead of a one-time or persistent pseudonymous buyer
16 address, pseudonymous buyer telephone numbers may be used for the third party to reach
17 the buyer under his/her terms. Example applications include:

- 18 (i) Financial Advice and Financial Planning Services. Often buyers are quite
19 sensitive about the confidentiality of the release of this type of information
20 related to personal financial matters and particularly with certain matters (and
21 perhaps in general) prefer that their financial advisors were unaware of their
22 true identities. Similarly, investment advice or sales communications by stock
23 brokers are another application where similar buyer information is typically
24 disclosed.
 - 25
26 (ii) Insurance Agents & Brokers. For many types of insurance, (e.g. health, life,
27 casualty) personally sensitive information is disclosed by buyers to their
28 agents and brokers. Initially, before insurance services are purchased, it is
29 possible that useful detailed quotes and/or insurance advice could be provided
30 to a buyer pseudonymously.
- 31

numeric values associated with various medical tests, results for which are a numeric value. This data may be of relevance to pharmaceutical companies, alternative medicine vendor and clinics insurance companies hospitals physicians, clinics and home health care providers, the latter three of which may wish to advertise to patient prospects and extend their medical practices. The privacy architecture herein provided is a critical component for enabling access to user data by these commercial entities and is perhaps critical to the extent that users must explicitly authorize the transfer of medical records from either the physical copying or electronic duplication of an associated transfer of such information to an intermediary (SDI) which is trusted by the user.

(xiii) Medical information, such as medical conditions, medical history, active prescriptions, drug reactions, family history, possibly even genetic pre-dispositions (from a genetic profile). Medical insurance information may also be potentially useful for a prospective qualified accessory to be able to readily access in case of an emergency.

(xiv) Physical location information—Users or advertisers could, for example: a) Query a pseudonymous user database to access profiles that are in close physical proximity and match certain criteria, e.g. live in a certain geographical region, had recently attended a meeting or event (or is planning to attend a particular event) had recently communicated with a friend or associate. In another variation, a user could for example, submit a query pertaining to every user in a particular physical space, e. g., a room, hotel or convention center, e. g., identify all users present here who attended Internet World, 1995.

1.3.1 Buyer Infomediary

The central SDI data warehouse can be used by vendors that provide services of *buyer infomediaries*, for example providing buyers in B2C e-commerce applications with historical information about previous purchases of users. At present companies such as www.priceline.com make profits because many individuals post take-it-or-leave-it buy offers for goods that are above the reservation price that vendors are prepared to accept for a service. An information infomediary, based around information submitted by users about previous successful or unsuccessful bids would provide for a more efficient marketplace.

1 A buyer infomediary system can be built as an application of SDI to providing bid prices for
2 agents in electronic marketplaces which are agent-mediated and allow dynamic pricing, for
3 example the nascent marketplaces of on-line auctions, www.ebay.com and on-line reverse
4 auctions, www.priceline.com.

5 **1.3.2 Seller infomediary**

6 On the flip-side, the system of SDI can also be used to support third-party infomediaries for
7 sellers, that make their business by performing data mining on the information in the shared
8 data warehouse, and can sell that information onto other vendors. It is possible to identify
9 useful pricing models, based on historic purchase information from many different
10 individuals.

11 The system of SDI can also be used as a confidential database for the purposes of generating
12 statistics from sensitive data. For example, as a trusted system, manufacturers might be
13 willing to provide information about their productivity, margins, retainment rates,
14 production efficiencies, yields etc. The central SDI server could generate statistics, globally
15 for the manufacturing sector, and then individually for each manufacturer—as it relates to
16 the information provided by other companies. Similarly, it would be possible to use such a
17 system to compare salaries across different universities. While an individual university
18 might be reluctant to reveal information about its pay-scales to other universities, in the
19 aggregate this information is not sensitive—and a survey on salary can be useful to both
20 employers and job candidates. SDI is used to securely calculate statistics, without revealing
21 any information that might compromise the privacy of a single employer.

22 **1.3.3 Personalization Application.**

23 Personalizing information on-the-fly requires that a vendor has a data model, for example
24 that clusters its current user-base according to what they are likely to be interested in. We
25 can allow vendors to execute queries, for example clustering algorithms on profile
26 information, that use all the information submitted to the data base by users without
27 violating privacy policies of the owners of that information. This can be done by returning a
28 *classifier*, that will place new profiles into relevant clusters, and provide information about
29 the likely interests of a particular user. Given this classifier, we describe in Section 3,
30 “client-side data mining”, a method that allows the vendor to compute the results of a
31 personalization algorithm on the profile of an agent without receiving that profile. Simpler

1 experimented with such that optimal click-through or transactions occur particularly for key
2 pages or purchasables. Adding to this statistical data SDI user profile information is able
3 to reveal (via data mining) much more robust relationships between pages as these
4 relationships change with user profile features. Accordingly appropriate rules may be
5 provided for individuals based upon this general user profile information which is gleaned
6 from interests and behavior before visiting the site, i.e., either from the user-centric or
7 vendor-centric SDI (while the user visits the vendor's site or alternatively accesses an
8 affiliate site).

9 In another variation of SDI we can highlight content on web pages with information that
10 might be relevant to a user, even when the source of the web page is not personalized. This
11 can be done via collaborative filtering techniques, which might bring in feedback and
12 comments from other similar users within SDI that are stored in the central SDI database.

13 The iamworthit (user side) SDI database in cooperation with the vendor centric SDI
14 service can sell to the vendor centric SDI service or other industry or market research
15 organizations strategic information about the comprehensive behavior activities and user
16 profiles of visitors and customers of these vendors (as is suggested earlier in the spec).
17 Additionally, targeted survey questions may be presented on behalf of these entities in
18 order to extract further information which may be correlated with certain features and
19 attributes of these users. One such method, rapid profiling is detailed in the parent
20 application [INSERT US PATENT NO].
21

22 Vendors can utilize SDI's data mining interface to observe content and product
23 consumption affinities based on user profiles describing Web wide behavior. This
24 interface may also enable vendors to observe comparative click through and/or
25 transaction rates of their competitors, how these criteria are affected by user profile
26 attributes including geographic criteria (if relevant to that vendor) as well as how these
27 criteria, (including even individual multivendor customers) are affected by various types
28 of recommender and rule based engines and further the particular rules and weighted
29 feature correlations used in generating these recommendations (again as described above,
30 these rules and feature relationships may be derived automatically through a core sample
31 of users, who are iamworthit subscribers). A vendor can use statistical data to tune

1 certain features and implement certain rules, either directly or with the help of a user
2 centric SDI analyst. Moreover, direct real time response feed back may be collected on
3 these manipulations to the data model in iterative fashion.

4 **1.3.5 Example: Ad networks.**

5
6 The system of SDI provides a method for an advertising network which allows users to
7 receive personalized adverts that are relevant to their lives, and also allows vendors to boost
8 profits by placing one-to-one adverts instead of the traditional broadcast nature of
9 advertising in media such as television and newspapers. Each user can receive adverts that
10 are tuned to his/her profile, but without losing control over the information in that profile. In
11 fact, a user can configure price-rules to receive payment for receiving a well-targeted advert.
12 The trusted secure data interchange can operate as an “ad network”, allowing for the
13 placement of well-focused banner ads to market goods that are relevant to users of a
14 particular content site. Electronic banner ads provide the potential for one-to-one marketing,
15 when the advertising agency has information about the user that has just hit a site, together
16 with information about what the user is doing local to a site. For example a car manufacturer
17 is able to place a focused advertisement to a user that has just performed a search for new
18 cars in a search engine, to a user that is known to have a large family and a high disposable
19 income.

20 Ads can be targeted to iamworthit users based upon their user profiles. Ad networks can
21 use the profile information in SDI to target ads and reach their target customers. Ads can
22 be targeted with statistical data across multiple vendors that are subscribed to SDI.

23 Statistical data can also be used to determine estimated revenues for a vendor if she/he
24 were to interoperate with iamworthit and subscribe to SDI.

25 SDI might develop a decision tree to decide how to assign adverts to users with a particular
26 profile. Within SDI we can track users across multiple domains, to allow fine level control
27 over ads that the user sees.

28 There are two possible business models. Firstly, an Internet content provider could purchase
29 access to information placed by vendors and users within the Secure Data Interchange
30 database. This information may be “rented” for a period of time, and then whenever a user
31 visits the site of the content provider (possibly through the pseudonymous proxy server), the

1 the preferred embodiment, any inserted ads delivered by iamworthit can be “turned off” by
2 the buyer voluntarily unless the in kind value which the buyer receives in exchange for the
3 advertising mandates a certain minimum ad impression delivery. Also in light of the above
4 buyer concerns in the preferred embodiment, there are no additional impressions added to
5 the buyer’s web browsing experience.

6 Ad blocking technology can be deployed to block existing banners and replace them with
7 iamworthit ads. Ad blocking is commonly available, and the techniques used are well
8 known in the art. The HTML source for the ad banners may be either called-up from the
9 remote server (such as with an ad network) or alternatively from an ad server (typically
10 purchased by the vendor) on the vendor’s own host server. In either case, the ad blocking
11 software typically recognize the HTML source which originates from the ad server. A
12 directory of the HTML source for the various ad servers is maintained with the ad blocking
13 software (and presumably updated e.g. if new servers are added or change their IP
14 addresses). The location/physical dimensions of the portion of the page occupied by the
15 HTML source which the software removes (its “footprint”), may be readily replaced by
16 another replacement banner (by iamworthit) from wherever the proxy server resides
17 (typically on the client but potentially on the network). As a result of a potential difficulty
18 in which the ad server deliberately changes its IP address to avoid recognition, which occurs
19 constantly and dynamically, it may be possible to recognize portions of the page which
20 contain image “features” of ad banners, which are inserted from another HTML source
21 where that HTML source is linked to known vendor sites (which are also linked similarly
22 from other site’s ad servers and/or are known via their HTML sources to have previously
23 utilized ad server technology and a “new” HTML source appears instead, etc.

24 This model may be extended to other media domains e.g. replacing digital TV commercials
25 instead with targeted ads delivered and precached via iamworthit for insertion at appropriate
26 times i.e. during commercial breaks during standard video programming. Another variation
27 could be tailored to pre-loading iamworthit advertising to automobiles for insertion in place
28 of traditional radio commercials. In accordance with the parent patent application, (and as
29 suggested above), email may be a useful targeted ad delivery medium as well. Per the
30 patent case, the system is able to classify email (according to its source and content using
31 implicit or explicit actions of the buyer). Based upon certain desirable confidence threshold

1 settings, the system may automatically delete “span” and replace it with targeted messages
2 provided by iamworthit which the buyer is paid to receive pseudonymous physical mail is
3 yet another potential source of targeted mail for which the buyer may be paid in conjunction
4 with the user’s voluntary removal from direct mail marketing services which iamworthit
5 could provide in conjunction. iamworthit’s buyer compensated targeted advertising model
6 could be extended to a pseudonymous telemarketing service (as suggested above) using one-
7 time call session pseudonymous telephone numbers for telemarketers to reach desired
8 buyers. Finally, human sales persons could be the “ad delivery medium” for iamworthit.

9 **1.3.6 Example: Dynamic personalization of links within a web site**

10

11 Wide user profile data can also enable the dynamic personalization of links within a web
12 site, to create a virtual shop floor to match the predicted preferences of each user. For a new
13 user, that has never before visited the site, it is very advantageous for the site to already
14 know about the preferences of that user in order to personalize the goods and services that it
15 offers. The information provided at the secure data interchange, and gathered from the
16 transactions of a user with another vendor, is vital for this type of personalization to first-
17 time users. For return customers, an Internet site can also leverage **its** is information that it
18 has collected from previous interactions with the user, information that is collected locally to
19 the site.

20 **1.3.7 Personalized Search Engines**

21 The present state of the art for search systems involves the use of an extremely static
22 interface, which is not personalized for a user. In the previous patented disclosure
23 [INSERT U.S. PATENT NO.] we identify features of a user and anticipate areas of
24 content likely to be of interest. A personalized portal interface can then allow a user to
25 view categories that are presorted in terms of expected value to a user, *based around*
26 *more than the search term just entered.*

27

28 For example, if AltaVista knows that I am a research student, then the search engine can
29 bias results in the direction of information that is likely to be appropriate to my interests,
30 and away from other items. The personalization and filtering can be done based on results
31 that I have accessed in the past, and also on the basis of information that is stored about a

1 user in the shared SDI data warehouse. It is possible to greatly increase the relevance of
2 the search results through eliminating and prioritizing the potential range of results from
3 the query.

4
5 Within the system of SDI, where more information is readily available about a user, we
6 might now also look at what a user is currently doing, his/her professional expertise,
7 skills, proficiencies, etc., the reason for the user asking the question, for what purpose
8 was the question asked, for what purpose will the results be applied (if relevant), what is
9 the geographic and/or time period content of the query, what type of content does a user
10 prefer, e.g. humorous, wisdom/council, scientific, political, business oriented.

11
12 In this way we can begin to incorporate contextual content into search. We might also
13 **allow** users to contact other users who are determined to be experts in particular areas
14 browsed, with this information dynamically discovered by queries into the SDI data
15 warehouse.

16 Collaborative filtering can help to identify individuals with a close affinity to a particular
17 site/vendor, **and** community of interest or interest domain such, so that they can be
18 informed of opportunities to act as experts and receive referrals on demand. This added
19 ability to bring new content more efficiently and timely could be very valuable.

20
21 Another simple extension is to provide personalization to the results from a standard search-
22 engine. For example, we could use a user's profile to perform additional filtering of the
23 results of the search to refine the pages returned on the basis of their profiles and the user's
24 profile, i.e. using collaborative-filtering/clustering style techniques.

25 **1.3.8 Smart Browsing: Personalized Annotations, Filtered** 26 **Information.**

27 We can also introduce personalized annotations to pages that are displayed to a user, with
28 annotations selected to reflect the interests of the user, for example as provided by special-
29 interest groups sharing affiliated profiles, or as provided by other users. A couple of
30 **definitions** for this section: *exemplar* - the profile of target object or (as pertinent to

1 following description), user profile which is “most like” the profile of the cluster to which it
2 belongs, perhaps a median metric.

3 The Platform for Privacy Preferences (P3P) [CR 98; RC 99] provides for the ability to
4 utilize XML meta-tags to annotate Web pages, and within a system such as SDI we can
5 allow comments from previous visitors to a page to add annotations. Users can receive
6 annotations from the SDI data warehouse that are associated with annotations provided by
7 other users with similar profiles. One of the divisional applications of the parent case
8 “System for Customized Electronic Identification of Desirable Objects” [Herz 98b] relating
9 to the automatic creation of virtual communities suggests that users may be automatically
10 assigned to particular communities (e.g. chat groups, forums, etc.) for this purpose.

11 We might allow users to rate the annotations in the pages, allow SDI to learn how useful
12 annotations provided by certain users are to new users, and classify users as “experts” that
13 receive priority in the position of their current and future comments. Future comments from
14 users with a poor rating history for a particular content cluster may be deleted. A persistent
15 interface feature on the tool bar or side bar can provide for annotations to also be accessed
16 by users selecting certain profile features of users as they browse from page to page; for
17 example a user could identify the comments of a news article about abortion by users who
18 are self identified as advocates of the Women’s Rights Movement, ultra conservative senior
19 citizens, teen women or those with a strong interest in alternative medicine or the Catholic
20 Church.

21 The parent case [Herz 98b] further suggests that users may actively provide ratings in a
22 completely privacy protected manner according to various criteria of pages they browse.
23 Profile-based clustering of these ratings and annotations can allow a user to submit as a
24 query a user profile, to receive a page that rates or annotates a page, and a listing of highly
25 relevant pages to a particular type of user. This could also be used in a “reverse engineering”
26 sense, for example requesting the exemplar user profiles for users that visited and gave high
27 ratings to a particular site, and then using those profiles to find similar sites of interest, or
28 finding the user features that are in the exemplar profile to understand something about the
29 content of a web page. The browser interface may automatically display the related links
30 that are determined dynamically to be most relevant to a user, again statistically estimated

1 via the data in SDI of users' browsing habits, and/or via active page recommendations or as
2 book marks by those users as being of particular relevance or similarity to the present page.
3 Browsing methods may include search-based browsing, and also browsing via a hierarchical
4 navigation menu system, with users classified according to their behavior patterns and/or
5 ratings which have been actively submitted. The parent case [Herz 98b] also suggests the
6 use of hierarchical clustering for products, to help in a smart shopping system. Similar
7 products can be presented together where the feature criteria for creation of the hierarchical
8 cluster tree could be price or other criteria. In addition to allowing users to view item
9 selections according to desired selection criteria, a comparison shopping function can also
10 allow a user to view the attributes of buyers that tend to buy certain items to help to add
11 confidence to a user that he/she is getting the right product.

12 The above description also describes the use of a hierarchical menu through which groups of
13 users may be identified by their profile features (wherein a profile feature could even be a
14 rating criteria itself of for example, an opinion via a site survey). These features could be
15 used to either selectively filter-out content which falls outside of that criteria as the user
16 navigates the information or identify if/when pages encountered where these user rating
17 features are present, thus displaying this user statistical information in conjunction with the
18 ratings statistics and/or associated annotations if desired.

19 We can allow the user to use one or more organizations to provide features and annotations
20 to help in a search and information interpretation process, to add appropriate filter and bias
21 to information presentation as a user browses the Web. Similarly, a user might adopt the
22 endorsements of a friend for annotations, and editorial content about particular pages.

23
24 The availability of feature information about individuals that provide high ratings for
25 particular Web pages can also be useful for vendors that wish to find appropriate
26 locations for their advertisements. For example, given an "ideal profile" of a consumer
27 for a particular product, a web page/type of information can be determined that will be
28 useful to find appropriate customers. This is an alternative method to providing adverts to
29 particular individuals, instead choosing to provide adverts to limited information domains
30 that attract appropriate users. Vendors may request additional information, such as site-
31 specific page view correlations (including time spent viewing each page) in accordance

1 with user features. Exemplar user profiles and attributes of those users are certainly of
2 interest to vendors to which those sites belong as well as affiliate sites on which their
3 advertisements and/or syndicated products are advertised and sold remotely.

4
5 We can also profile users with context information, for example based on their assumed
6 goals (social, business, personal, etc.), information which can be determined by the nature of
7 their current activities.

8 In an interesting extension we can allow users to contact other users that are browsing
9 similar material with similar profiles in real time, allowing ad-hoc virtual work group
10 formation. Of course, all of this is done within the carefully managed profile and privacy
11 managed systems of SDI. The location of a user could play into decisions about physical
12 meetings. Typically some users may wish to enter into a present (or future) dialogue, which
13 can be scheduled via calendar agents, which perform automatically scheduled compatibility
14 meeting/introduction functions.

15 It is also reasonable to provide the technique of collaborative filtering to identify users
16 whose profiles are particularly similar to the material which is being viewed, e.g., those
17 users who are determined to have a high proficiency level with regards to that particular
18 material. Because the requesting user may be seeking to query the expert or seek edification
19 on that material, typically same consideration is conveyed in exchange for that information.

20 In a variation, a more ----- data exchange may be provided if collaborative filtering
21 identifies fairly similar users to the presently viewed material which also demonstrate, via
22 their profiles, considerable complementarity in the particular knowledge which they are
23 likely to possess. Vendors may also ----- and facilitate this type of
24 information exchange on their sites between these complementary, (yet metrically similar)
25 customers.

26
27 In a consumer-vendor commerce application of this form of interactive browsing we can
28 allow users to enter into dialogues with vendors based on their profiles, and share
29 experiences with vendors, regarding their experiences. This can be done in real-time, or
30 statically within an iambworthit style system architecture, where a user receives financial
31 incentives for providing information that is useful to vendors.

1 The parent case [Herz 98b] suggests that *organizations (e.g. vendors, etc)* may be rated by
2 multiple users across and according to a variety of relevant attributes. One useful extension
3 is to group descriptive opinions of users according to their profiles, certainly users with a
4 particular type of profile are likely to have affinities towards certain sites and certain
5 content. It can be useful to example how the assessment of user's varies according on their
6 profile information.

7 Another variation is to suggest a *three-dimensional interface* to allow useful diagrammatic
8 representations of the information and meta-information that a user is presented with. The
9 parent case [Herz 98b] describes a method to reduce the multidimensional characteristics of
10 feature and annotation information to only 3 dimensions, and present that information
11 graphically, via a technique of principle-component analysis. The profiles of the target
12 objects associated with a user's preferences are represented as points in 3-D space within a
13 navigational interface. Nodes may also be displayed according to category and sub-category.
14 Views may also be restricted to user's with the same goals, i.e. business users, social users,
15 etc. this mind-set can be inferred by the sites that a user is visiting, for example.

16 **1.3.9 Smart Interface to On-line Discussions.**

17
18 The present description provides a methodology by which it is possible to apply techniques
19 of statistical NLP for purposes of enabling users of live discussion groups and bulletin
20 boards to dynamically visualize concept (via automatically generated labels short summaries
21 or even visual representations of the concepts). The parent patent [Herz 98b] describes a
22 browsing implementation in which it is possible for a user to view an on-line 3-D layout of a
23 "virtual mall" in which the multi-dimensional space of the object profiles of the items are
24 collapsed into three dimensions. Further, the display may be customized to the user's
25 profile by using as the vantage point the position of the user's profile within the space.

26 We can use information retrieval, pattern matching, and collaborative filtering techniques to
27 refine information based upon the particular relevance feedback from those particular
28 individual engaged in a discussion. The novel interface enables the extraction of text
29 segments from all previous discussions. These (multi-discussion) segments are essentially
30 grouped into clusters and the clusters are representative of "concepts". The statistical
31 framework is further used to determine what clusters (concepts) tend to follow in sequence

1 is ever released. That said, within the iamworthit system an individual with a current and
2 complete political profile could receive a steady income stream from semi-automated focus
3 group queries.

4 Taken one step further, it could then also be possible to identify for a user candidates which
5 are believed to be a close match with a user's political profile. Expert opinions could be used
6 to determine the individual issues and relative weightings (indicating relative priority which
7 those issues assume within the politicians political priority scheme) as well as levels of
8 extremism in views to political issues. In the preferred implementation a listing of different
9 politicians is presented in conjunction with a percentage which reflects the degree of
10 similarity among the views of the user and those of each politician as listed in descending
11 order or similarity. The user could drill down and observe a redistribution of the percentage
12 based on particular desired political issues, or the user could observe the distribution under a
13 variety of different user profiles. The system may also take aggregate statistics from other,
14 various types of individuals and or organizations. That is to say each individual can, on a
15 voluntary basis, disclose as much personal information, be it explicit or implicit disclosure
16 (form his/her user profile data base), in order to create a more robust and comprehensive set
17 of statistical relationships.

18 The system could also ask the users presently to indicate which actual candidates they
19 selected; thus collaborative filtering recommendations can indicate, for example, which
20 candidates users with certain political issues-related profiles tend to vote for (even what
21 other individuals, exactly like the present individuals, tend to do in this regard), or
22 conversely, what political issues voters of certain particular candidates tend to claim to
23 subscribe to. The system may also identify what characteristics (political or other wise) are
24 most prevalent or alternatively most unique to individuals which share an affinity for a
25 particular candidate or a particular issue, or the system may recommend candidates based
26 on existing candidate affinities.

27 The techniques of the parent patent [Herz 98b] may also allow a synchronous and
28 asynchronous discussion with other individuals containing certain profiles (or identical
29 profile characteristics). Again the pseudonymous server may be useful in this regard. The
30 present system may also enable rating of candidates, in which various types of individuals
31 (in aggregate) rate the candidate according to different criteria, characterizing the candidates

1 on issues such as strengths or weaknesses. This is a useful polling example, the ratings may
2 be revealed according to various types of individuals.

3 **1.3.13 Example: A “Virtual” Jury**

4
5 The judicial process **attempts** to be as fair and equitable a possible by allowing a jury of
6 “peers” to be entrusted with the responsibility of evaluating the guilt or innocence of a
7 particular individual as well as to determine the degree of liability, the penalty and degree
8 of financial compensation which justified in light of the circumstances. The information
9 within the system of SDI could be used to store information about users, explicit and/or
10 implicit information based on their browsing habits, and offer a service, e.g. in civil cases
11 for jury selection based on characteristics of individuals. In a criminal case the “virtual jury”
12 may be used in addition to a real jury, to provide additional feedback on a case. The virtual
13 jury may be paid by the plaintiff to participate, e.g., in the case. The feedback from the
14 virtual jury may be presented to the judge and/or an actual jury before its own deliberations
15 and verdict. In a novel variation the virtual jury may even hedge bets upon its conclusions
16 such that if they contradict the conclusions of the actual jury but are later determined to be
17 correct the members receive compensation. The advent of live streaming video greatly
18 facilitates the feasibility of the current methodology.

19 In a variation, it may be possible to apply the techniques of collaborative filtering towards
20 automating the selection instead of those individuals which are representative of the most
21 diverse possible groups and clusters of individuals a cross a wide range of criteria (including
22 such criteria as psychographics, ethnicity, intelligence, social status, etc.) In fact, subject to
23 some changes in the law, this approach would be a more efficient means of selecting a
24 standard jury than the process of random selection.

25

26 In one variation, the virtual ----- is selected based upon similarity to the plaintiff. In
27 another, it may ----- of individuals which are individually similar to both plaintiff and
28 dependent on more effectively a combination of individuals who are similar individually to
29 the plaintiff and the defendant respectively. It is believed, in addition to guilt innocence, a
30 more equitable assessment of the monetary compensation which **is provided to the plaintiff**

1 by defendant , it such a equal mixture of empathizers from both sides constitute a
2 virtual jury.

3 **1.3.14 Example: Group Therapy**

4
5 There is a very useful and appropriate application for SDI to the automatic matching of
6 individuals for purposes of group therapy. Individuals with commonality across many
7 criteria can be selected, and multiple long-term groups adopted. In an application of this
8 system, it may be possible to broadly extend the present scheme to the general public (where
9 privacy is secured via the proxy server) and individuals may identify a group(s) which best
10 fits their own unique emotional needs. It may also be useful to archive the sessions, index
11 enabling the sessions and segments thereof to be searchable by keyword, e.g., via speech to
12 text techniques and/or browsable by topical segment (which may be automatically
13 segmented and labeled). The present system may also provide an ideal framework by which
14 psychologists may identify patients of other psychologists that similar pseudonymous
15 complements of psychological characteristics and symptoms, to allow targeted clinician
16 interactions and robust and useful information from therapeutic approaches and/or drug
17 treatment regimens.

18 In a variation, the present system methodology also enables a means by which much more
19 specialized group session topics may be created which focus upon a very specific type of
20 disorder, conflict or aspect of the patient's psyche. In this way, it may be possible for
21 "identical" patients to, as group, focus upon different aspects of their ideal therapeutic
22 regimen which collectively create a comprehensive customized treatment program. It is
23 worth noting that the specialization possibly associated with particular focused sessions (and
24 even types of individuals) provides a framework by which clinicians can become extremely
25 specialized and expert within certain specific sub-domains of the field.

26
27 It is certainly possible to adapt the above described technique for "smart interface" to
28 provide useful ideas and suggestions for patients engaging in either psychotherapy or group
29 therapy thus leveraging the information and personal experiences within the therapeutic
30 processes of the collection of patients precoding them within similar topical sessions which
31 shared very similar psychological and pathological patent profiles.

1
2 It is certainly reasonable and appropriate to adopt the present application framework to more
3 several fields of medicine. For example, enabling physicians to identify other physicians
4 have ----- had patients with the identical pathological profile and medical history.
5 This enables them to ----- current medical history and clinical data, insights,
6 observations, etc., in order for the physician to ----- the present clinical based on the
7 expert advice and collaborative feedback from the other physicians possessing very similar
8 experiences. Likewise patients could instantly access the physician or physicians which
9 have or had experienced the most similar clinical situations that and the present situation of
10 the patient in order to seek a second opinion. Physicians sharing unique clinical experiences
11 (particularly clinically or scientifically in----- or series/problematic could use the -----
12 ----- scheme in order to ----- ----- companies, etc.. Finally one could also apply
13 the -----
14 -----
15 -----.

17 **1.3.15 Example: A Personalized Educational Portal**

18
19 Another interesting application is to develop and deploy a personalized on-line
20 informational portal containing everything from helpful links, illustrative content, text book
21 information, quiz questions etc. The Personalized Educational Portal (PEP) typically is
22 designed specifically for a particular class and includes several key features which are
23 optimally applied as part of a comprehensive intelligent educational system. These features
24 include the following:

- 25 • The ability to dynamically customize all forms of relevant information from the
26 educational portal. The criteria for this customization however is based not upon
27 the preferences of the user (unless for example the user receives credit for studies or
28 projects or research on topics which s/he may select, rather it is based upon a
29 predicted profile of the user reflecting his/her strengths and in understanding the
30 relevant content. In particular the techniques of the issued patent number 6,029,195
31 [WHAT IS THIS PATENT NAME] describes a variation of user profiling in which

1 and if a wrong response is submitted, the system may recommend reading certain content
2 which has (as above suggested) been statistically demonstrated to improve the student's
3 proficiency in that sub-domain. It may also be possible to gauge in advance the students'
4 predicted level and rate of advancement by knowledge domain via on-line psychological
5 testing. Statistical correlations between these test questions and the student's learning
6 abilities (by knowledge domain) can be usefully developed.

7 The system could even be constructed hierarchically by graduation of skill levels i.e. initial
8 mix of questions of varying difficulty levels in order to gauge what level of content to
9 recommend. The questions which follow are one level higher as is the next set of
10 recommended content unless improvement is not observed.

11 It may be possible to a decision tree which automatically select which questions most
12 effectively test the students, command of the materials (in this way a shorter exam may
13 actually be better than for longer exams those questions which are more exemplary of such
14 knowledge may carry greater weight). We could also effectively try to create a summary
15 e.g. by topics /headings or even using text summarization techniques submit the areas of
16 deficiency in one student to another student who is particularly proficient in that area for
17 virtual tutorial. For this the tutor can receive monetary and/or also scholastic credit. The
18 ideal scenario is actually creating virtual study sessions in which students are matched
19 together which have the most different complement of proficiencies/deficiencies as possible
20 and where these areas of potential mutual exchange benefit are maximized and are revealed
21 to both students initially. Other forms of virtual study groups could be achieved by
22 grouping students by similar proficiency areas and levels and creating a "virtual tour" with
23 questions and content presented synchronously to a group of correspondence enabled
24 students. In this way, if virtual tutoring or study groups occurs on-line exclusively, it may
25 be performed pseudonymously as desired. We can even take a defined study group and
26 assign it to group projects which can (again) relate to content to which the group is
27 (collectively) deficient. (as group oriented problem solving has been proven to be extremely
28 effective in remedying such deficiencies.

29 In a very novel application, it may be possible to even statistically using historical data to
30 analyze the relative proficiency profiles of students by knowledge domain (perhaps) even
31 further by, learning ability in those domains by teacher. Thus a virtual class could be

1 constructed using live streaming video which are truly customized to each student's
2 particular needs and abilities, i.e. the idea would be before the semester to identify which
3 teacher(s) were most successful in eliciting the highest levels of achievement among
4 students which have an identical profile to that of the student (such statistics used in this
5 profile may include but are not limited to the proficiencies / grades of the student in all other
6 previous classes, psychological testing and/or (selective) knowledge domain proficiency
7 questions (as above described).

8
9 Certainly depending upon such investor's personal preference, it is reasonable for each to
10 agree within their personal data disclosure policies to disclose to SDI data regarding the
11 types of investments, amounts and under what conditions (including those conditions stated
12 by other investors such as the example above). The conditions for this disclosure, however,
13 would be that no individual data be disclosed to another investor, i.e., that it be used only for
14 SDI to be able to reveal aggregate statistics and predictions how the investment community,
15 in aggregate, interacts with investment opportunities, and under what conditions, and in
16 exchange for this disclosure, each investor earns the right to access this aggregate data. The
17 individual investor or SDI, acting on behalf of that investor, may then suggest the best terms
18 and conditions with which to negotiate with other investors, in order to optimally achieve
19 their own desired objectives, part of the condition statement to SDI could also include the
20 amount of money to invest in each synergistic investment opportunity which SDI typically
21 would suggest to the investor based upon the relative distribution of its presently invested
22 funds, size of the available non-invested funds as well as the relative risk/benefit proposition
23 of the company to that investor (as estimated by SDI's use of company data as disclosed to
24 that investor). SDI can even recommend or act as a proxy, in negotiating with the individual
25 companies (who themselves could use an SDI negotiating agent representing their own
26 interests). E.g., SDI, acting for the investor, could leverage considerable information about
27 not only, the present investment opportunity, but literally all other investment opportunities
28 which had been submitted to SDI for consideration by SDI affiliated investors. Based upon
29 the total benefit which the investor stands to gain which is based, in turn, upon the SDI
30 recommended comprehensive investment strategy) and the relative downside (relative risks

1 compared to other alternative prospects within the total pool of investment opportunities)
2 may be used to determine basic parameters in which the investor or its SDI proxy would be
3 willing to deal, e.g., the amount of funding, the associated amount of equity as well as other
4 more subjective factors such as amount and nature of control, the use of the funds (e.g. will
5 they be used for a particular technology product or service which, if developed, offers the
6 primary degree of benefit to the investor's existing company(s). It is also worth noting that
7 SDI may identify those prospective investors whose collective cooperation would result in
8 an economically optimal scenario for that investor. SDI, acting on behalf of the collective
9 group of investors, may recommend data disclosure terms to each individual investor (or its
10 SDI proxy) which will optimize the probability that the coalition resulting from the
11 cooperation of all or most of the recommended participating investors will optimize the
12 economic benefit to that individual investor. This data could be no more than predicted
13 value (including marginal value) that particular investor would achieve the desired coalition
14 were to be formed. Based on this data the investors or their SDI proxies could negotiate
15 with one another in order for each investor to strive towards its own individuals best interest
16 while attempting to assure that the global objectives of the coalition are ultimately achieved.
17 E.g., that not only the best investment opportunities within this optimal global plan are
18 funded but also the least desirable as well (of course this information would be concealed
19 from the companies at least until the negotiations are concluded). Accordingly, the relevant
20 investors may wish to compete with one another for the best investment opportunities within
21 this global plan or SDI with their unanimous cooperation again acting in their collective best
22 interests may as a coalition assign investment opportunities to each investor assuring them
23 each equitable treatment and, of course,...no competitive bidding for a given investment
24 opportunity (which, of course, would result in ultimately more favorable terms to the
25 collective group.

26

27 In this preferred application (of non-competitive bidding) the SDI agent for the collective
28 group achieves a commitment from each prospective investor in the group that within a
29 range of prospective terms, each investor would be willing to commit its participation in the
30 proposed investment opportunity. Accordingly, each company (assuming there are more
31 than one) also agrees to a predetermined range of terms with which it would be willing to

1 accept from the collective group of investors. Both sets of information (typically
2 recommended by the SDI representation each associated party) are disclosed to the main
3 (party neutral) SDI entity. If overlap exists in the range of terms or the ranges are close to
4 one another, the introduction and negotiation between the parties becomes initiated (through
5 the company(s) are not made aware of whether the overlap exists or is merely “within the
6 general proximity” of accept-----
7 -----introductions on the interest of the parties (though most markedly the
8 companies)
9

10 **1.3.16 Example: A “Group-Think” Ideas Market**

11
12 In the consumer driven market place there is a significant unrecognized opportunity in
13 being able to harness the intellectual capital of the consumer market. Historically, (in a
14 non-networked environment) leveraging of such knowledge has been impractical and
15 virtually infeasible due to the inability to identify and measure this resource of human
16 skills (and thus match this skill with in the appropriate matching problem)with the
17 particular types of problems which companies may be in need of new ideas and solutions
18 as well as the lack of a trusted intermediary which can validate on an impartial including
19 basis the authentic contributions of the provider of the information (as not having been
20 claimed by the recipient after the fact as having been previously conceived before). Of
21 course, there are remaining obstacles such as the fact that important problems are
22 sometimes maintained confidential within a company (not even disclosed to some its own
23 internal staff).
24

25 With the emerging of extranets a certain amount of inter-organizational knowledge
26 leveraging is becoming more readily achievable. A secondary advantage over an extranet
27 is because of its ability to act as a trusted intermediary. It can validate the fact that that
28 the individual is only dealing with one commercial entity and not its competitors (at least
29 formally). Thus unlike extranet based knowledge leveraging it is conceivable a trust
30 relationship enabling a certain amount of privileged information disclosure is at least
31 theoretically possible.

1
2 In the preferred embodiment, SDI retains a record as to the particular problem/solution
3 sets that were successfully or unsuccessfully provided by users in the past in order to
4 improve predictions as to their abilities to successfully address the same types of
5 problems/needs in the future as well as to more accurately determine the “similarity” of
6 different and (perhaps often apparently) disparate problems from the standpoint of the
7 user’s ability to solve problems based upon knowledge of their ability to successfully (or
8 unsuccessfully) solve others SDI is entrusted with the knowledge of particular needs of
9 organizations (it may, for example, cooperate with Think Task companies). These needs
10 are commonly oriented ideas (though not exclusively so), for example, suggestions or
11 ideas for product development, marketing, a business model or concept (such as a new
12 business direction), content creation, technology or even potential commercial customers
13 or strategic partners. In lieu of identifying particular problems to which an organization
14 may be desiring a solution, users may instead provide their own ideas and/or solutions to
15 needs or problems which they perceive SDI vis-a-vie its in-depth knowledge of the
16 internal needs and problems of numerous entities is thus able to determine the relevance
17 of the suggested solutions or ideas or alternatively identify a particular organization
18 which is likely to possess matching need which is relevant.
19

20 **1.3.17 Experimental Validation of New Models: A Test-Market**

21

22 The system of Secure Data Interchange can be used for a *test market* application, where
23 the effect of a vendor’s new collaborative-filtering based targeting model, or new pricing
24 model, or advertising campaign, can be tested by simulation and “hold-out” testing.
25 A small sample of users can be targeted, and the results measured and reported to a vendor,
26 before a vendor commits to an extensive campaign. This may provide a cheap method to
27 test market conditions for a product, and run tests on sample groups selected on the basis of
28 similarity to other products for which buyer response is known. Random sampling of
29 buyers may be initially performed using data mining as applied for marketing research.
30 Based upon this initial response vendors may identify key features of respondents which are
31 typical buyers, non-buyers, proponents, non-proponents, etc. Accordingly, rapid profiling

(as similarly) may enable the automated or semi-automated selection of disclosed segments of users (containing particular attributes) which are likely to have relevant properties. The user selection process may be iterative based upon responses from users as they are selected. We can demonstrate this value experimentally, for example we can offer a vendor a free-trial and present personalized information/advertisements to one group of SDI users (both on the site and if desired also across the network), and regular advertisements etc. to another group without the aid of SDI. The increase in vendor revenue can be estimated from client-level monitoring of the change in purchase volume achieved with well-focused solicitations **on the vendor's own business**. Other prospective ads and syndicated purchasables could likewise be virtually overlaid on the vendor's site. Thus degree of increased click through rates, transaction rates and syndication revenues could be tabulated for that vendor in advance of his/her subscribing (or even being solicited) to SDI. Portals are also potential beneficiaries of this service in light of their need to better target their partner's ads. The ability to the tailor a targeted marketing strategy to users knowing these correlations could be extremely beneficial in both on-line as well as off-line commercial environments. In the process of selecting the target user profiles significant value may be achieved from data collected from that vendor's own on-line presence, and a portion can be sold to other similar vendors; as discussed in the main description of SDI, this now forms a cached query which can be used by other vendors.

We might also demonstrate value to vendors with SDI by monitoring the performance of vendors with SDI technology, and providing metrics for new vendors to allow them to select suitable models of user targeting and personalization. Vendors that subscribe to SDI (the vendor centric version) can provide more attractive offers/products to users, based on information about the wider activities/interests of a user, on other vendor pages, and in the physical world (of course, only to the extent that this information is authorized by the user). Vendors can use information in the central SDI data warehouse for users' cross vendor and within vendor browsing and purchasing habits, and also with respect to profiling information about a user.

One key application is first-time personalization, so that information and products can be targeted to a user when he/she first hits a web page, based on profile information that the

1 As the trend to disintermediation continues (whereby we have informational middlemen, but
2 not middlemen that physically hold products) we can allow vendors to notify portal
3 intermediaries of product location within SDI, so that the portals can then pass this
4 information onto customers.

5 The key role of the SDI system in this application is to allow users to receive personalized
6 recommendations of products, and also product locations, based on their profile
7 information—but without the vendors receiving that profile information. In addition, the
8 shared information about user profiles in the data warehouse allows vendors to build good
9 models for inventory location. Finally, SDI can act as a trusted intermediary between
10 different competing vendors: for example suppose Amazon users bookstore A and
11 bookstore B to provide books, and A and B do not wish to tell each other where their
12 products are. The stores tell SDI where the books are, and SDI presents that information to
13 users on a per-book basis, as books are browsed within the purchasing system.

14 There are numerous useful applications to improving quality, speed and cost of delivery to a
15 user. In one example, it may be possible to provide same day delivery for on-line purchases.
16 We could also restock a truck on the basis of what a local population of users are likely to
17 purchase; with items presented to a user along with a particular anticipated delivery period,
18 with the Global Positioning System (GPS) or more LEIA [INSERT THE PATENT NO.]
19 on the vehicle providing up-to-date information about a truck's location and anticipated
20 delivery time.

21 We might also suggest that a user can physically travel to some location close to his/her
22 base, based on information within SDI about the user's travel patterns, etc. Purchasable
23 items can then be located in locations that are convenient for users, or shipped dynamically
24 to those locations (i.e. intermediate warehousing systems). Users can be provided with real-
25 time directions to the location of such a pick-up point via personal digital assistants (PDA)
26 and navigational devices, etc.

27 1.4 Push-based Advertising/Solicitation

28 Another form of information that a user can submit to the central SDI database via his/her
29 client is the advertising acceptance functions, which state the explicit preferences of a user
30 for adverts, and required payments. The level of compensation that a user requires to receive
31 an advertisement will depend on the relevance of an advert.

1 An application of SDI is that vendors can use the centrally stored information to request that
2 as a result of a query the query-execution module in the central SDI data warehouse sends
3 advertising solicitations to appropriate users, i.e. users that are willing to receive adverts as
4 specified within their acceptance functions, and willing to release profile information to
5 enable useful advertising as specified in their price rules for regular data.

6 Figure 15 illustrates the system for push-based advertising. A vendor requests the right to
7 advertise to users by sending a special type of query to the central SDI data base, where the
8 query states the vendors preferences for user profiles and requests that adverts be sent to
9 users. The vendor never receives the contact information for user's directly. The server also
10 receives payments from vendors and credits users' accounts.

11 Suppose Vendor V wants to advertise a new product to relevant users. Vendor V can use
12 the SDI database to select users with a high predicted hit rate for the advert, and then
13 compute the value of a bid that it is prepared to pay users for the right to provide them with
14 its advert. The vendor makes a request for bids to appropriate users by formulating this as a
15 query with an associated action, and submits to the central SDI data warehouse where it is
16 handled by the query-execution module. Part of the query is a bid function, indicating the
17 maximum amount that a vendor will pay to display an advert.

18 The query-execution module evaluates the request with respect to the profiles of each user,
19 the rules that agents submit for allowing access to profile information, and the advert
20 acceptance functions. If successful the vendor pays the minimum value necessary to have its
21 ad accepted by the user, again simulating a Vickrey auction as for the dynamic competition
22 for banner ads described earlier.

23 For example, suppose a vendor has an advert **Ad** to push to users, and is willing to pay up to
24 \$1 to place the advert with users that have a predicted hit rate of greater than 30% on the
25 advert. The vendor can formulate a query that will first identify users with a predicted hit
26 rate that is greater than 30%, and then determine which users will accept the adverts at the
27 bid price. If successful, the adverts are pushed to the users via the SDI central data
28 warehouse, and the payment required by each user is submitted, less than \$1 in all cases. In
29 a simple variation, the vendor might also state a fixed budget, so that it does not send
30 advertisements to more users than it can afford.

31

1 We can use the same profile information that provides focused/personalized service to users
2 that hit a site that they have not visited before to form well-targeted mailing lists for
3 vendors. The Secure Data Interchange can form mailing lists in a number of different ways.
4 First, consider a vendor that wishes to send targeted mail to some of its own user-base.
5 When users connect to a site they indicate whether or not they are willing to receive
6 electronic mail, and provide a "mail certificate" to a vendor if they are happy to receive
7 mail. The Secure Data Interchange can proceed as follows:

- 8 (a) Perform analysis for the vendor to determine an appropriate set of users to receive
9 the solicitation, based on the information that the vendor provides about what it
10 intends to market, and provide the list of pseudonyms to the vendor for mailing;
- 11 (b) Perform the same analysis, but also forward the communication to the users
12 directly.

13
14 Now, consider a vendor that wishes to target new users, represented with different
15 pseudonyms. Users indicate whether the information that a vendor submits about his/her
16 transactions may be used for solicitations, and furthermore vendors indicate the set of
17 business interests that can receive the benefit of information that is submitted to the central
18 SDI server. The SDI server can continue by performing analysis on the relevant subset of
19 the permitted class of data records that pertains to the product or service that the vendor
20 wishes to model, and generate a list of appropriate pseudonyms. Finally, the SDI server can
21 sell the pseudonyms to the vendor outright, together with a certificate that the vendor can
22 send mail to the pseudonyms, or the SDI server can retain control by sending the mail on
23 behalf of the vendor.

24 Provide vendors with virtual mailing lists that can be mailed to via the proxy server only. i.e.
25 these customers should be solicited based on our analysis. (could even give summary info.,
26 without revealing details about users). Furthermore, do not even reveal data that corresponds
27 to a pseudonym to a vendor because the vendor then has that information about me when I
28 am on his site.

29 A central data warehouse also enables vendors to identify new potential customers. This
30 process is broken down into a number of steps:

- 00699098 " 10E" 00
- 1 (i) The vendor assesses the value of the information present in the secure data
2 interchange. This computation is performed securely either by revealing randomized
3 aggregates to the vendor to enable its own local analysis, or by allowing the vendor to
4 check data and algorithms into the secure data interchange site for analysis.
- 5 (ii) The vendor selects criteria for mailing unsolicited advertisements, and agrees on a
6 pricing model. In this case per-impression pricing is the most obvious pricing model,
7 as it is difficult to monitor when a user responds to unsolicited mail per-transaction
8 pricing is difficult. The user could be motivated to do this should the Secure Data
9 Interchange promise future returns for recording a successful solicitation with the
10 database.
- 11 (iii) Either the data list is released to the vendor for its use, if this is within the selling
12 vendor's data policy, or the data interchange sends mailings on behalf of the
13 purchasing vendor.

14 1.5 Community Dollars and Business Models

15
16 The primary objective of the iamworthit model is to create a market for information
17 about buyers, with agents that submit profile and other personal information to the
18 database able to collect payment in return for queries executed by vendors. In an
19 important extension, we can allow this payment to be made in terms of *community*
20 *dollars*, which can only be spent with particular vendors. Community dollars allow a user
21 that provides profile information to the system of SDI to receive payments that are
22 *dedicated for a particular type of purchase*, these payments are called "community
23 dollars". One central example is to allow a vendor that signs a user into the SDI system to
24 be able to lock a proportion of payments accrued by that user to the vendor's own
25 product/service domain. This is important, because vendors now have incentives to bring
26 new users into the system. A vendor can offer a user community dollars on its (and its
27 affiliates) web site, in exchange for receiving user data via the SDI data exchange.

28
29 We allow vendors to pay in "community dollars" for adverts, dollars that can only be spent
30 at that vendor (with the host site of the advert receiving a share of the profits). This provides
31 vendors with the ability to gain long-term customers. Furthermore, so long as the buyer

1 agrees to receive advertising from his/her iamworthit subscription offer, community dollars
2 can be replenished at the rate at which advertisers are willing to pay for impressions. This
3 provides buyers with an incentive to spend at the vendor's site, because the vendor can
4 monitor (pseudonymously) the buyer's that are sensitive to discounts and other special
5 offers (that are delivered as community dollars). With community dollars a vendor can
6 compensate buyers for information that they access, but tie that information to certain
7 vendors (e.g. the vendors that first signed the user into the system). The system has the
8 following useful properties:

9 (1) buyers are incentivized to provide information that allows vendors to push relevant
10 advertisements/products;

11 (2) buyers will also be more likely to make purchases at a site or a coalition of sites for
12 which they can receive discounts via community dollars;

13 (3) providing buyers with community dollars will increase the number of hits to a site.

14 We allow buyers to receive compensation for providing personal data to vendors,
15 information that has value to vendors because it allows information to be focused (for
16 example relevant ads can be displayed to a buyer, based on his/her profile). The system of
17 iamworthit credits buyers for information, and provides buyers with direct incentives to
18 reveal profile information to vendors.

19 A vendor can sign up with iamworthit.com and agree to provide only the most restrictive
20 type of community dollars, that can be spent at that vendors site. Community dollars are the
21 currency that vendors provide in return for the right to provide focused information to
22 buyers. Dollars can be general (e.g. for a network of vendors), or very tightly focused (e.g.
23 for a particular product, at a particular time). The buyer-centric infomediary acts as a broker,
24 matching buyers and vendors. Another key role of the infomediary (e.g. the portal) is to
25 protect the buyer from information saturation by controlling the flow of solicitations. (i.e.
26 restrict the number of ads. that a buyer sees)

27 Community dollars extend the methods in Secure Data Interchange that allow a user agent
28 to sell controlled access to information to other agents, introducing in addition a method
29 which can control the ability of a user to spend received payments. For example, a vendor
30 might be happy to pay a user for the right to display a targeted advertisement if the payment
31 represents a discount that can only be used against products offered by that vendor.

impressions per day). Given these significant benefits, a vendor can provide a buyer with a significant discount (in the form of community dollars). Vendors benefit from increased sales volumes. When the price of items is less than the value of dollars, the vendor can limit the amount of discount that is available on any single product, or only allow community dollars to be applied towards customer discounts (which may nevertheless be quite substantial).

1.5.2 Implementation Details

We can use meta-tags to restrict the way that community dollars can be spent. The tag associates the dollar, but the dollar is released within the system of blinded signatures [Chaum 85; Chaum 92] so that a buyer that collects dollars over many transactions with different vendors can spend the dollars without compromising his/her private information about pseudonyms. Dollars can be restricted to a number of vendors, and also restricted in additional ways—i.e. they can only be sent if the buyer visits the site through a particular portal, cannot be redeemed at a competitor, are worth a bonus if redeemed with certain vendors, etc.

The technology used to implement community dollars is based on cryptographic techniques known in the art, for example within a digital signature scheme for electronic-cash as described in the work of Chaum [Chaum 85; Chaum 92]. The core functionality which must be supported by any system for community dollars within SDI is:

- a) An agent can receive community dollars under multiple pseudonyms but spend the dollars under the same pseudonym. The dollars must not allow an adversary to link the pseudonyms and break a user's identity.
- b) Community dollars can be restricted to be spent with certain vendors and for particular products/services.
- c) Community dollars can accrue (and lose) value in non-linear ways.

The function of (b) can be supported within a system of electronic coupons, based on cryptography, as disclosed by Tygar et. al. [Tygar 98]. Alternatively, we can extend Chaum's [Chaum 85; Chaum 92] technique of blinded signatures to allow dollars to be

1 which is possible via the one-way identity look-up functionality of the central SDI data
2 warehouse. The eBank can query the look-up table, and check that both PIDs correspond to
3 the same UUID.

4 For example, an amount of dollars below a certain threshold can be programmed to have no
5 value, but additional dollars can have increasing marginal value. This provide a non-linear
6 incentive for a user to stay with the same vendor in a market with low switching costs, or a
7 non-linear incentive for a user to receive adverts from the same advertising agent. The
8 advantages to vendors and advertisers are repeat purchases, consumer lock-in, and also
9 improved profiling from interacting with the same agent over extended periods of time.

10 The functionality is embedded into the community dollars. The framework expands the idea
11 presented in Chaum [Chaum 85; Chaum 92] where the public key of a public key /private
12 key pair indicates the dollar value of a community dollar. In this case, we allow any number
13 of public-key/private-key pairs, and use the public key as an index into a table maintained
14 within the eBank to provide additional criteria about how the cash can be redeemed, for
15 example it has more value if used in combination with other community dollars, it decays
16 over time from data YY/YY/YY, etc... We do not intend to limit the variations possible, but
17 provide this as a framework for ways in which constraints and conditions on community
18 dollars can be extended. For example, community dollars can also be programmed to lose
19 value over time. This can increase the number of repeat purchases

20 In the preferred implementation we use an “electronic cash” infrastructure for the
21 community dollar system. A buyer’s SDI-enabled client-level proxy stores dollars that the
22 buyer receives securely. Dollars are anonymous and non-traceable, so that the buyer can
23 maintain a single “bank” of dollars, and aggregate dollars collected across pseudonyms for a
24 single purchase, so long as the purchase satisfies the constraints on the dollars. Each dollar is
25 created using Chaum’s blinded signature technique, and also signed with the conditions on
26 its use.

27 This scheme allows vendors to monitor the offers that buyers respond to, because when a
28 buyer presents a community dollar—the dollar can be validated to indicate the type of
29 discount that it is, even if the identity of the dollar (i.e. the serial number) is untraceable. SDI
30 provides vendors with guarantees that buyers have once-in-a-lifetime pseudonyms, so
31 redeeming a voucher of a particular type that is redeemable only at vendor V and was issued

by vendor V allows vendor V to be sure that the voucher was issued under the same pseudonym, and has not been transferred to another of the buyer's pseudonyms.

In an alternative architecture, the ISP-level SDI proxy, or the web-host for the advertising service, can maintain community dollar "debit" accounts for each buyer. This is more limited, because it does not allow buyers to transfer dollars between pseudonyms without compromising privacy (revealing a portfolio of pseudonyms). However, in a scheme where advertisers require that agents have once-in-a-lifetime pseudonyms, and only release community dollars to be redeemed at their own site, this is not limiting. Both of these approaches are useful for "community dollar-enabling" numerous or all sites.

1.5.3 Example Community Dollar Applications

- On-line Gaming site.

Consider an on-line gaming site that has a network of affiliated vendors, that do not pay to advertise, but provide community dollars that can be spent either at the vendors—or at the casino. When buyers lose money at the casino the casino receives real dollars from vendors. The casino is one possible outlet for spending dollars—and a vendor only pays the casino (the host of its ads) if a buyer chooses to gamble on the site, and loses its dollars. The gaming site becomes a portal, with links to partner vendors. Each vendor offers the buyer community dollars, that can only be spent back at that site or at the casino. However, the number of dollars which can be spent at the store is substantially less than the number of community dollars which can be spent at the casino. IF the buyer accepts the credits, whenever the buyer accesses the URL to the store he/she is either automatically routed first to the casino portal or to the vendor site whereby a prominent banner is displayed which is displayed to that particular buyer from which the buyer can conveniently engage in a casino gaming session.

If the buyer loses a substantial amount of community dollars he/she may regain the lost credits by spending a specified amount (in real dollars) at a partner vendor. This provides a safety-net for buyers. The cost to the vendor is the cost of the dollars that

the buyer lost at the casino, and the cost of replenishing the buyer's community dollars (which can be used for further gambling). However, the vendor makes a sale—so the vendor is happy so long as the dollar value is a reasonable discount for the sale. Clearly, the casino gains substantially through the redemption of these community dollars. The vendors can make an agreement with the casino where they only compensate a fraction of community dollars. A percentage of a buyer's profits at a casino are paid in community dollars, another percentage can be paid in real dollars. Perhaps the casino can also provide vendors with a revenue share.

We allow for community dollars that are restricted to particular products, and customized for an individual buyer. The dollar-object can contain two parts: the first part is readable to the buyer, and indicates the nature and the amount of the discounts to which the credits can be applied. The second part of the message is encrypted, and accessible only to the vendor, and is signed by the vendor to prevent any form of alteration. The information can contain the dollar credit to the buyer, the terms and conditions of the community dollars, a dollar amount, the pseudonym ID of the buyer, an expiration date, the terms and conditions of discounts and special offers to which community dollars may be applied in combination with a partial cash transaction.

The vendor must check that it has not previously redeemed any piece of community currency with the same identifier; the identity of the buyer is correct; the date; the terms and conditions. Some vendors may allow redemption of the community dollars at other vendors' sites.

- Web-hosting Application.

One of the commercial applications of which community dollars may be used is in offering web sites free web development services and/or free web hosting services. In the former case a Web development tool provider could share in the community dollars revenue stream from iamworthit ad revenues, and accordingly cover the costs of the tool suite. Whenever new upgrades including new development modules are added, the costs could be covered by an incremental increase in the percentage of advertiser

in the parent patent application) iamworthit, user referral links may additionally personalized based upon their user profiles.

1.5.4 Vendor coalitions

Vendors may choose to form coalitions, to allow buyers to spend community dollars at any “partner” site. Vendors that have similar buyer bases can be automatically identified using collaborative filtering. (i.e. determining similarity with the present vendor, from the aggregate vendor preferences of a given vendor’s subscribers). Also, these resulting metrics could incorporate predicted online spending by each buyer at each site. This could help to narrow the selection of sites the vendor wishes to partner with and/or the selection of these partner sites could be determined and presented to the buyer to even further narrow the selection for each buyer. All vendors in a coalition advertise, and provide cross-links and up-links to other vendors.

The coalition model is good for buyers that are more likely to find products that they want. Vendors can share the risk of advertising, since dollars provided to one buyer by a particular vendor can be redeemed at another vendor. Advertising and community dollars increases sales volume at all vendors in the coalition. Furthermore, studies on on-line buying behavior suggest that on-line shoppers tend to make purchases across a variety of categories of e-commerce products, thus it is likely that they would also prefer the freedom and flexibility which is associated and can be provided in this way only by deploying a multi-site community dollars scheme. Iamworthit usage statistics are very effective in identifying prospective vendor coalitions which consist of complementary (non-competitive) vendors. Alternatively, personalized coalitions which can be created in ad-hoc fashion for each buyer can further serve the buyer’s interests by increasing the freedom and flexibility which the buyer often strongly desires. Though practically this can only occur for those (perhaps smaller less commercially significant) vendors which have not been a major coalition consisting of vendors which many buyers tend to use which collectively serve a substantial percentage of the overall user base.

Thus if the vendors have not established firm partnerships with other vendors, we can even allow vendors to form dynamic and virtual coalitions within SDI, with a potentially unique coalition of vendors for each buyer. The coalition may consist of an optimal pool

1 of vendors, as determined by SDI collaborative filtering techniques. The goal in this
2 model is to provide buyers with a particular “brand” of community dollars.

3 • Multi-vendor Community Dollar Portal

4 We can allow each vendor to retain an exclusive right to advertise to each buyer; and
5 also develop a portal for the coalition—that gives advertising prominence to coalition
6 members. Portals will be expected to aggressively promote community dollars.

7 Buyers that collect community dollars become loyal return visitors to the portal and
8 its associated vendors. In the case the vendors do not generate the same value we can
9 provide community dollars in proportion to the value that a vendor contributes to a
10 coalition.

11 We can also provide targeted advertisements for the vendors at the portal, using the
12 user profile to focus ads. The categories and links at a portal (that might include a
13 search engine) can be re-prioritized (highlighted and/or re-ranked) in accordance with
14 the buyer’s preferences (as described above), and to favor subscribing vendors. In
15 commercial practice, the motivation for vendors to become coalition members is
16 largely driven by customer partnerships with providers of e-commerce tools and
17 solutions. There are many emerging trends by which these intermediary tool
18 providers could conceivably integrate community dollars. Consider loyalty points
19 (e.g. www.mypoints.com) or loyalty bonuses programs (e.g. www.clickrewards.com).

20 These points or bonuses could be substantially subsidized by the advertising and
21 deeper benefits passed onto the consumer. Other simpler technology which would
22 compel these vendors to cooperate include shopping basket technology, the emerging
23 standard, ECML, common Web tool and/or tool development solutions, common
24 hosting solutions, common ad delivery systems.

25 Vendors pay the portal site to advertise, and the portal provides community dollars to
26 buyers in return for privacy-protected profile information. This model does not
27 provide incentives for the portal to provide well-targeted adverts, because there is no
28 direct link between a portal’s revenue stream and the vendors’ sales volumes.

29 A portal with community dollars that can only be spent under a single pseudonym at
30 its partner sites also provides an incentive to buyers to interact under a single
31 pseudonym—which in turn allows a portal to profile buyers across its complete

1 vendor partner network. Buyers will access many sites with the same pseudonym.
2 The system of SDI allows vendors to leverage the shared profile information as
3 buyers browse web pages and products.

4 We can also lock buyers into a single portal—and a single coalition of vendors—with
5 community dollars that “decay” over time, and must be continually replenished. In
6 this way a buyer cannot pick and choose different portals, and different community
7 dollars, but can benefit mainly from high web-browsing volume through a single
8 portal. The value to vendors in terms of consumer lock-in can be considerable.

- 9 • Affiliate Network. A coalition of vendors can join to allow a buyer unlimited
10 access over all affiliate vendors. The program can be sold through the existing
11 marketing channels of each vendor, as well as through a portal directory of sites
12 for those vendors. Vendors that join can be required to promote the program
13 through their own marketing channels. Additionally, perhaps vendors are selected
14 to cover exclusive physical regions (e.g. in the case of a set of ski resorts), or
15 exclusive product categories (e.g. in the case of on-line vendors). Vendors can
16 provide a community dollar-for-real dollar exchange, in return for becoming part
17 of a vendor network. Alternatively, perhaps vendors provide an up-front fee,
18 that can be recovered via dollars spent by buyers at their own site. Each vendor
19 is obligated to sell the partner network community dollars, but is not necessarily
20 required to promote the other community dollar vendors.

21 An affiliate program could be developed by partnering with the affiliate network’s
22 advertising. Typically each advertiser provides a purchasable(s) which affiliate sites
23 offer for sale on their sites in exchange for a percentage of the resulting sales. In the
24 community dollars variation, each vendor agrees to tailor an offer for their
25 products/services on the affiliate network which is subsidized by community dollars
26 which the user collects after subscribing to iamworthit. An example solicitation
27 could be “receive \$100.00 free credit toward future purchases at booksamillion.com
28 plus \$50.00 cash.

- 29 • Creation of an iamworthit online multi-store retail site. Establish a retail presence
30 in a (or potentially multiple) retail niches. The primary business model would be
31 to leverage existing large iamworthit subscriber base (involving the other various

types of commercial partners) in order to dedicate a certain percentage of the community dollars (e.g. thirty percent or approximately one hundred fifty dollars per customer) which could only be redeemed at that multi-store retail site (and/or the value of these dollars could be worth more at the retail site). In addition, in this model, the independent advertising initiative of iamworthit would be geared towards community dollar credit of that retail site. It should be noted that, because if other outside competition occurs to the basic iamworthit scheme to a substantial degree there will not be a compelling incentive for buyers to adopt a more restricted form of value (as retail credits at a particular site), versus accepting the credit from a competitor in the form of cash. Thus this model could provide a viable means for attaining a leading position in one or more on-line retail markets if this competition does not substantially exist.

- **Marketing Network.** Iamworthit sites which offer a community dollars promotion could, upon the buyers subscribing to iamworthit, additionally offer the buyer with a down-loadable client based software which provides a small promotion in conjunction with a link to iamworthit. Each time a recipient of the email subscribes to iamworthit, a percentage of the value of that customer is credited back to the buyer in the form of community dollars. Each subscriber resulting from the current subscriber's email (though reduced) provides an additional credit to the original subscriber in accordance with the marketing network business model. If the site originally delivering the promotion is not an e-commerce site, a percentage of advertising revenues resulting from the subscriber (and potentially all resulting subscribers) could be used. It could be applied in the form of iamworthit advertising (or exchanged) for advertising in an ad server.

- **Free Web hosting.** A portion of community dollars may be allocated in a revenue sharing arrangement between the buyer and the web-site from which subscriptions to the service are provided, thus enabling Web hosts to offer free hosting services while also receiving full payment which is iamworthit advertising supported. Iamworthit could also easily use its vast statistical data collected from each site in

1 We can also require that buyers are automatically routed through a portal when accessing
2 any partner vendor directly. The portal (and therefore the coalition of vendors) then
3 receives exposure each time the buyer clicks on an ad (or link) to that vendor. The portal
4 may also provide benefits (e.g. additional advertising prominence) for sites which are
5 responsible for driving traffic through the portal. Community dollars can be provided
6 whenever the buyer accesses a site from the portal. Portals can offer free advertising to e-
7 commerce sites (forgoing advertising fees). The portal provides discounts to buyers that
8 purchase a product following a link provided at the portal.

9 A buyer receives the discount by validating a purchase with the portal, and the site agrees
10 to provide the portal with a share of revenue whenever the buyer cashes in community
11 dollars in this way (we do not rely on HTTP refer mechanism because that can be
12 blocked and falsified. Furthermore, we do not rely on URL+extension correspondences,
13 also not secure—instead rely on providing buyer's with incentives, and monitoring
14 buyers that have earned community dollars.)

15 **1.5.6 Delivering per-impression dollars**

16 We can deliver community dollars on a per-impression basis, with vendors competing to
17 offer buyers high values for being able to present an advert. The existing collaborative-
18 filtering engine at a buyer's SDI client-level proxy can filter ads, and select appropriate
19 offers, using community dollars as just another measure of the usefulness of a message.

20 This is an alternative to providing dollars on a one-off (or even yearly) basis, for
21 consumption via the vendor's site that the buyer subscribes to the service.

22 A hosting site can take a fraction of any dollars provided to a buyer. Alternatively, a site
23 can convert the value into the community dollars to provide to the buyer, possibly at a
24 preferable basis. The portal might also wish to convert its commission to credits for the
25 buyer at any one of its partner vendors, with the stipulation that the buyer must access
26 those sites via the portal in order to be able to redeem the credits.

27 **1.5.7 Providing Loyalty Bonuses**

28 We can use the client-side SDI proxy to provide vendors with "loyalty guarantees", that are
29 credentials to verify that the buyer has executed no transactions with any competitor, under
30 any of its pseudonyms. The client-side SDI proxy is in a unique position to be able to

1 implement this monitoring, because no other system knows a buyer's portfolio of
2 pseudonyms. The buyer can present its digital credential when visiting a vendor's site.

3 A vendor may wish to provide loyalty dollar credit; for example, it would be possible for
4 vendors to offer buyer's credits if the buyer is a 100% loyal customer i.e. that she/he did not
5 (over a specified period) do his/her purchases at the site of any competitor. For example,
6 certain types of high value customers could be given considerable value in the form of
7 credits or discounts as a result of demonstrated vendor loyalty. The credential can be time-
8 stamped, to prove loyalty. It does not reveal any information about the buyer's other
9 pseudonyms to a vendor, because many pseudonyms will exist that have not made any
10 purchases from a competitor. Upon accessing the vendor's site, this credential may be
11 presented to the vendor. One criteria for the above benefits could be that the buyer may visit
12 a competitor site, and engage in interactions; however s/he should not transact with that
13 vendor.

14 1.6 Alternative Business Models

- 15
- 16 • Offer discounted or free services in return for the right to access profile
17 information.

18

19 Examples: Free or discounted retail products with "niche" partners in each category;
20 Free dial-up ISP (as an independent ISP or a service to jointly promote free access
21 with ISPs); Free Cable and ISP service, Free pay-per-view (note that viewing patterns
22 and the associated content could provide additional valuable user profile
23 information); Free phone service (e.g. advertise subscription service on screen phones
24 or audio ads from pay phones); Free prepaid calling card; Free print media
25 subscriptions (magazines, newspapers); Free book clubs; Offer any combination of
26 the above with "deep discounts" for each (this can involve \$350 community dollars
27 per buyer or it may simply involve certain purchasing limitations per customer).

28 Each vertical niche partner in exchange gets exclusivity within their own respective
29 niches to target advertise to those buyers (e.g. retailers); Free access to sporting
30 events; Free credit for casinos; Free lottery tickets; Free charity donations;
31 Discounted hotel lodging; Monetary credit to a credit or debit card (either an

1 iamworthit branded card or provided as a partnership with the card companies;
2 Monetary credit to a diner's club; Free subscriptions plus credit to retail buyer's clubs
3 (on-line or off-line); Credit or discounts for book clubs; Free musical concerts, or
4 theater presentations, movies or access to arcade entertainment; Free access to
5 amusement parts or theme parks; Free golf season passes; Free commission fees for
6 stock trading; Free commission fees for travel booking (if implemented for on-line
7 buyers would be less compelled to search for travel information on-line though go
8 off-line to make their bookings).

- 9
- 10 • Allow the ISP to promote free Internet access through personalized advertising on
- 11 iamworthit.
- 12

13 The ad server can even recognize through the associated domain names, the buyers
14 which are coming from a competitor ISP. So long as that ISP is not a partner of
15 iamworthit, the associated buyer would be selectively targeted with an offer of this
16 sort "free Internet access" by subscribing to iamworthit". Smaller ISP's would be
17 particularly compelled by such offers to their direct target prospects. This is because
18 they are operating on a "thin margin". Furthermore, both they and their small
19 regional counterparts would be particularly vulnerable to this type of advertising by
20 regional competitors from the same geographical area, during specified period of
21 months of initial usage of the service, the share of profit due iamworthit could instead
22 be committed to purchase additional advertising for the Internet service provider (or
23 the balance of this profit traded-out in the form of additional advertising through the
24 ad server partner).

25

26 The ad server partner could further become an exclusive partner of iamworthit on the
27 following commercial venture: Relationships as established with on-line merchants
28 and other e-commerce sites. The vendor actively promote an offer to their customers
29 through both off-line media (using a URL unique to that vendor) and on-line
30 advertising through the ad delivery partner. The offer may say (as an
31 example),"receive three hundred dollars in purchasing credit at Books a Million in

1 exchange for subscribing to iamworthit (or receive five hundred dollars worth of
2 discount credits at Books a Million. In addition, as a further benefit to the vendor
3 iamworthit could trade its own advertising impressions with impressions on the ad
4 server for the current offer (in order to reach a different base of buyers which are not
5 currently subscribed). A particularly compelling industry for this application is on
6 line travel inasmuch as a plaguing problem to this industry is the fact that many
7 buyers use travel sites as an informational resource on available travel deals and
8 packages, but ultimately book their trips directly through a travel agent (thus cutting
9 out the travel site). A three hundred dollar a year travel credit would be a compelling
10 incentive to many buyers to modify their current travel booking habits. Affiliate
11 networks are also an ideal channel for these types of promotions because affiliate
12 sites agree to participate (typically) purely based upon the degree of the profit sharing
13 opportunity (which would be significantly larger than most types of transaction -
14 based affiliate advertising).

15

16 Alternatively, it would be possible to offer websites the ability to become Internet
17 service providers where the interface to the ISP home page would essentially be
18 heavily branded to that site or portal. Companies like GTE already offer a "Virtual
19 ISP" service in which the content to the ISP home page is unique to the ISP while the
20 network is provided by the virtual ISP service. This model would be particularly
21 compelling for sites which are largely community oriented and have a potentially
22 loyal customer base. Moreover, interestingly, many of these community sites are
23 offering many of the services and capabilities that a full-blown ISP would offer from
24 its home-page, e. g., a portal interface, links to high-quality content, chat/forums, e-
25 commerce, commerce affiliate links, etc.

26

- 27 • Free Community-based Content.

28

29 Create premium content which is free to iamworthit subscribers, subsidized by
30 revenue from profile information. Each iamworthit buyer would be granted free
31 access privileges to the premium content on all sites which are part of the program.

- In addition to the aforementioned free community site content, it would be possible to further provide free and automatic access to fee-based information on the Web. These costs may be able to be covered by the model across all or most sites depending upon the usage characteristics of its buyers (e.g. assuming advertiser/community dollars payment to the sites are averaged across buyers in accordance with the consumption patterns of the average ianworthit buyer). The identity (pseudonymous) of the buyer would have to either be disclosed to the site via the proxy or a unique pass code (as required by the site) provided to the buyer could be automatically entered upon the buyer accessing the fee for use area requiring the code. A directory (portal) of these fee-based sites would be a useful adjunct to subscribers.

- Patent Application - SDI.doc

1
2 Many PC manufacturers are now recognizing e-commerce as a very important sales
3 channel. The present model would involve the P.C. manufacturer bundling a link
4 along with a promotion for iamworthit. The promotion would offer the buyer cash
5 credit for the buyer. The PC manufacturer would also receive exclusive advertising
6 rights to target buyers whose browsing behavior profile qualifies them as a future
7 sales prospect. E.g., instead of cash credit as the profile is identified, the offer could
8 then become modified to offer free hardware or credit to their purchase. Because PCs
9 are highly portable, the advertising targeting techniques described in LEIA could add
10 substantial additional value to advertisers. A similar model could be used for
11 manufacturers of PDAs.

- 12
13 • Allow matchmaking of buyers based on profiles across a network of
14 iamworthit/community-dollar sites.

15
16 A menu of different forums and chats can be displayed on each iamworthit member
17 site. (the underlying methodology for which is described in co-pending patent
18 application "Virtual Community Service for System for Customized Electronic
19 Identification of Desirable Objects"). In accordance with this specification, a
20 variation of the service involves the process for identifying individuals who most
21 closely match a given category or target object. For example in the context of the
22 present implementation a category or content, merchandise or a purchasable being
23 specially promoted may be the focal point of a discussion forum or chat room, which
24 is automatically organized by the Virtual Community agent. Accordingly, a portal (or
25 in accordance with the present trend) a site with which a portal interface is integrated
26 utilize the present techniques for generating virtual communities for each category or
27 sub-category of content on the portal or for direct access into a forum or chat room
28 which was automatically created around that particular site (as the target object used
29 as the matching criterion). As described, the user may navigate a hierarchical menu
30 of virtual communities which may be constructed automatically according to the
31 methods described which involves communities assigned to category, sub-category,

1 and association with corresponding sites. Ideally in this scheme the portal is actually a
2 “virtual portal” which may be utilized in providing access to the communities across
3 numerous sites (and/or ISP home-pages). Users may also be navigated (at the
4 individual user level) which along with their pseudonymous user profile data is
5 subject to their data release policies. In a variation of the above schemes, if there is
6 geographical information which is associated and which is released in accordance
7 with the above individuals and/pr communities (e.g., as may be occurring or
8 scheduled to occur in physical space), LEIA may be employed as a primary (or
9 additional) selection criteria for navigating the present information accordingly.

- 10
11 • Advertising in Exchange for Equity.

12
13 A potentially attractive optional form of value, which could be provided to
14 iamworthit customers involves equity shares in companies which advertise to the
15 buyer (in lieu of community dollar credit or cash). This scheme is an ideal
16 application for iamworthit in as much as iamworthit customers can be highly targeted
17 and because many Internet-based start-ups are highly niche community oriented (thus
18 iamworthit customers who are interested in the sites can be efficiently identified and
19 targeted). Moreover advertising is typically very expensive which in the absence of
20 accurate targeting may be of questionable value. It should be noted, however, that
21 because the primary objective is to both find viable prospects and to engender an
22 element of loyalty (which the equity model does). This scheme would be the
23 preferred approach to advertising for sites which do not sell on-line where community
24 dollars would be the preferred loyalty engendering scheme. In order for this model to
25 substantially provide its desired advantages of increased advertising exposure to
26 fledgling web based companies, the iamworthit subscriber base would have to be
27 quite substantial.

- 28
29 • Loyalty credits for off line retailers.
- 30

1 Deliver through the back of sales receipts, kiosks or direct mail or on-line substantial
2 purchase credit to retailers (e.g. grocers') customers, using the aforementioned
3 technique of utilizing a unique URL to identify the vendor and/or promotion from
4 which an iamworthit subscriber originally accessed the iamworthit subscription site
5 (thus identifying for both buyer and vendor the appropriate denomination and/or
6 terms of community dollars issued to the buyer). In the preferred implementation, a
7 loyalty card is used to identify the buyer thus enabling the community dollars value to
8 be provided to the customer at check-out as straight credit or possibly an
9 enhancement to loyalty credit. The buyer may also be identified via credit card or
10 alternatively a voucher (or coupon) could be printed from the buyer's computer or
11 from a kiosk which is typically situated near the entrance to the store and which could
12 be activated upon insertion of a loyalty card credit card (or associated authorization
13 code) and could also be used to disclose the buyer's community credit balance. A
14 unique identifier for that voucher or coupon is provided and non-tamperability
15 measures are provided such that the buyer's community dollars account can be
16 appropriately debited upon redemption. Preferably, a pre-determined value is
17 specified on each voucher (which could be predetermined by the service or the buyer)
18 or alternatively, the total community dollars balance could be specified on the
19 voucher along with the buyer's name/address and redeemable only upon presentation
20 of valid buyer ID.

21
22 • Free ASP Services – Web-centric applications are becoming an increasing central
23 part of the e-business infrastructure. iamworthit could potentially enable
24 organizations to gain free access to these Web-centric applications in exchange
25 for iamworthit enabling their employees and customers (e.g. which may use some
26 of these applications as well).

27
28 • Credit to a User's Credit Card – Many credit cards are tailoring promotions to
29 enhance not only acquisition of market share but also loyalty of its users. The
30 user-centric SDI provides an enhanced (e.g. rebates for expenditures) platform
31 which could provide more efficacious loyalty enhancement marketing strategies

1 for card issuers. For example, a credit card user could be given certain credit
2 towards the card (e.g. via a promotional offer) for becoming an iamworthit
3 subscriber (which is largely ad revenue supported). An additional novel
4 application could involve providing an offer for an additional benefit which could
5 be provided if the user performs all of his/her on-line transactions using the
6 present credit and (i.e. wherever the card is accepted).

7
8 An iamworthit pseudonymous credit card [e.g. LMP 94] could be provided whereby
9 users collect credit for advertising (e. g., on billing statements for the ad) which the
10 user receives as well as the advertising delivered via iamworthit's on-line (Web or
11 targeted email) as well as pseudonymous physical mail, telemarketing calls using user
12 profile pseudonyms and one all time or pseudonymous phone numbers. This user
13 profile data consists in part of off-line data and on-line data.

14
15 • Free Personal Portal for Individuals.

16
17 It is reasonable to offer individuals completely free, Web design/development and
18 hosting services which are offered and mass marketed. At the time this patent was
19 written, Web developers were offering such free services in exchange for revenue
20 sharing for advertising product syndication or e-commerce offering on the
21 individual's site. Iamworthit could conceivably be deployed in conjunction with this
22 commercial model whereby either a portion of the user's own iamworthit profiles are
23 used to subsidize (or subsidize in part) the portal and the developer and/or revenues
24 shared from profits from iamworthit subscribers who subscribed from the portal are
25 utilized also (or instead). As such, it is also reasonable for such a developer to
26 provide links on affiliate sites (e.g. access an affiliate network) which offers free
27 individual portals in exchange for subscribing to iamworthit (and perhaps agreeing to
28 offer a similar iamworthit expectation from that portal as well from which the user
29 can also share in a portion of the revenue stream. Such a program could even be
30 structured in its revenue distribution to the subscribers as a multi-level marketing
31 network. Users could even receive value for providing links to the vendors site either

1 both (e.g. as “cob----- shopping portal) ----- direct off-line
2 encounters in the other users as prospective customers to the URL. (where the URL is
3 unique for each individual.
4

- 5 • Multi-level Marketing Network.
6

7 This economic model for distribution of shared revenues from iamworthit may be
8 implemented several ways. For example, as above suggested, the user’s Web portal
9 offering solicitations for iamworthit (as a standard offering or in combination with a
10 free personal portal), could be implemented in a number of different ways, such as ;
11 a). The solicitation is presented upon the user’s Web portal. b). The solicitation is
12 made through the traditional sales channel of direct one-o-one correspondence with
13 friends or associates. C). The solicitation is attached to any email messages of the
14 user such solicitation provides a link to enable the user to subscribe.
15

16 1.6.1 Agent-Mediated Value Exchange in the Supply-Chain

17
18 The information in the SDI data warehouse can allow vendors to identify business
19 relationships between other vendors, for example between vendors and their suppliers,
20 and between vendors and their customers. With this information a vendor might attempt
21 to exert pressure on another vendor by making direct offers and incentives to their
22 suppliers and/or their customers. Of course, in all cases information is only available
23 subject to price and disclosure policies, but with there being at least two parties in any
24 deal no vendor can unilaterally prevent this kind of information from reaching the SDI
25 enabled data market place.
26

27 Information may include the benefits/disincentives associated with particular actions, and
28 help to identify actionable events. Agents might offer other vendors a share in some
29 future profits that result from taking a particular action, so that little up-front collateral is
30 required. The information in the data market can be used to estimate the value of this

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

The mechanism for providing value to the recipient (in exchange for providing the desired action) is flexible and may include (as suggested above) “upside” value, direct compensation (or direct “redeemable value”), or an exchange of actions between agents (I will do X if you will do Y). Part of a deal may be that the provider of a requested action guarantees exclusivity and that it will not provide a similar action for competitors.

A vendor might identify and propose deals to the customers of another vendor, to apply economic pressure through threatened boycott and compel the vendor to perform a desired action. This “indirect” economic leverage may be applied at the consumer (lowest) level in the supply chain, and passed upward. In addition to boycott, a vendor could promise please note, shouldn’t consumers be termed customers of the vendor since consumers cannot be threatened with boycott) consumers credit in return for taking a particular action. SDI can mediate the entire electronic control, transfer, fulfillment and negotiation processes (typically XML-based) individually on behalf of each entity within this multi-step negotiation process. In the above applications SDI can represent each entity in the negotiation process, and also have a role in providing information and statistics, expert opinions, etc. which clearly elucidate the value and potential value opportunities available to the entity(s).

The current model allows different supply chains to compete for the loyalty of the consumer, e.g., through providing indirect economic value to the consumer as in kind products/services and/or upside in the benefiting entity.

Economic alliances can be created between vendors, suppliers and consumers within the same supply chain. SDI can act on behalf of different supply chains and establish a competing market between different chains for the business and ultimate loyalty of purchasers. End purchasers have a large economic leverage on a supply-chain. The decision of an end purchaser will depend on:

- 1) The type, selection and quality of personalized products/services

1 With interwoven supply-chains, for example, with the same supplier connected to
2 multiple vendors, and vendors with competing suppliers, this method of passing value
3 along the supply chain becomes more tricky. When a consumer applies economic
4 pressure on a vendor that it deals with in the supply chain, that vendor may inform other
5 suppliers of the pressure and block service to the user. A bartering system can provide
6 additional benefits, to allow vendors to exchange bonuses, discounts, etc. with favors
7 offered to vendors in other value chains. In general it is quite likely that a favor will not
8 substantially benefit a user in another supply chain. We can also include a method to
9 prevent a vendor that does not itself offer favors in the supply chain from taking
10 advantage of favors offered by other vendors. This can support cooperation
11 within the supply chain, and allow effective competition with other supply chains.

12
13 Each vendor in the supply chain can benefit in two ways:

- 14
- 15 1) They receive a “commission” on the transactions in which they convey the favor
16 to their customer (which again may be “upside” in the ultimate beneficiary of the
17 favor and/or product/service loyalty credit with their own supplier).
 - 18
 - 19 2) (Most importantly), they receive customer loyalty advantages via the community
20 credit they provide such customers.

21
22 For reasons of this latter benefit they are likely to compete with other competitive
23 vendors for the privilege of exclusively delivering the incentive to evoke the favor to the
24 next level up the supply chain. In this way each vendor below the supplier requesting
25 stands to gain via customer loyalty benefits by competing with another supply chain.

26
27 A consumer can provide value to a supply chain by exclusive provision of personal data,
28 to allow a supply chain to create specially customized and targeted products and services.
29 However, it is not necessary for a customer to have an exclusive contract with a supply
30 chain (although this might bring greater loyalty benefits and payments).

31

1 A variety of different techniques and methods should be integrated and deployed to
2 implement this agent-mediated supply chain value exchange system:

- 3
4 1) Digital contracts to define and enforce particular transactions and long term
5 business relationships.
6
- 7 2) An SDI infrastructure to suggest particular opportunities for a vendor, based on
8 the benefits to the complete supply chain. SDI can utilize as much information as
9 the vendor is willing to entrust in SDI, to try to identify present and predictive
10 future benefits resulting from actions on the part of other vendors.
11
- 12 3) A means for enabling customers and/or vendors to form coalitions, to strengthen
13 their ability to provide pressures on other vendors.
14
- 15 4) A mechanism with useful incentives for these customers and vendors guarantee
16 direct benefit even if the desired action ultimately does not result in new value,
17 e.g., an insurance scheme.
18
- 19 5) A method to exchange terms and conditions between members in a supply chain,
20 by being passed up the supply chain from the lowest level (consumer), and a
21 mechanism to provide “fair compensation” for this process.
22
- 23 6) A method to integrate loyalty benefits and long-term purchasing contracts, etc.
24 into an electronic cash protocol, so that buyers are aware of all contractual
25 constraints during negotiations with agents.
26
- 27 7) Methods within SDI to advise a customer about an appropriate bidding strategy,
28 based on estimated values of different contracts and supply chains, to help the
29 customer to make a good decision.
30

1 8) Methods within SDI to recommend information to reveal to a supply chain, to
2 maximize the value of a loyalty bonus offered and ultimately received by an
3 agent. Of course, information cannot be falsified, but some information can be
4 withheld.

5

6 Applications of “bribing” the consumers of a vendor:

7

8 • A commercial entity is sustaining bad publicity from an article published in
9 newspaper X. The company could offer the customer base of the publication
10 (identified via SDI) an incentive to temporarily disrupt or boycott the publication
11 until corrections/changes are made to the article. The company could also offer
12 discounted products or services to the consumers, based on information (from
13 SDI) about their consumption patterns. A more indirect discount could be
14 provided via other vendors further up the supply chain, via multi-vendor
15 exchange.

16

17 • Company A is a high tech start up selling software products and services to
18 established vendor B. Vendor B identifies the substantial value proposition of the
19 products and services provided by the start up, and decides that it should provide
20 the services for itself “in-house”. Start up company A can go to the consumers of
21 vendor B (information from SDI) and offer a percentage upside, e.g. percentage
22 equity in the start up, if they will apply a threat to boycott the vendor unless the
23 vendor agrees to do business with the startup. Company A might even identify
24 consumers or suppliers further down the supply chain, and apply indirect pressure
25 to B.

26

27 • Company A is developing a new commercial initiative which may be competitive
28 to that of another smaller company B, and might harm B’s market share. Assume
29 that consumers prefer B’s products to the future products of A, and that B can
30 make more profits than A. Company B could convince the customers of company
31 A of these facts, offering them a percentage of potential upside and/or discounted

1 or value credit towards products. The customer profiles of company As
2 customers also reveal the vendors from which they most frequently purchase, and
3 company B could also offer equity or otherwise some potential upside to these
4 vendors in exchange for offering discounts or credits to their customers (which
5 are customer A's customers as well).

6
7 In place of monetary compensation, we might provide a value payment to an agent in the
8 form of a long-term compensation, based on the long-term value achieved by an agent
9 further up the supply chain.

- 10
- 11 • SDI can identify the paths in the supply chain between companies, to allow a
12 sequential transfer of benefits down a chain. Ideally, each entity providing the
13 benefit is doing business with another next lowest in the supply chain, so that
14 value can be exchanged.
 - 15 • SDI might “bribe” each party within a value chain to enter into a contractual
16 relation which benefits that party (directly or indirectly).
 - 17 • SDI can determine what companies along this chain would have the greatest
18 future potential to benefit one another, i.e., where the supplier can generate
19 compelling long term benefit to each of his subsequent buyers at each level down
20 the chain.
 - 21 • SDI can use available information regarding a vendor's internal and external
22 activities and goals, to identify optimal multi-vendor supply chain models, in
23 terms of maximizing the sum value to all members in the supply chain, *and* the
24 value of each individual company.
 - 25 • SDI must also consider the willingness of each vendor to establish the necessary
26 contractual relationships for the new supply chain. Vendors might provide SDI
27 with “relationship policies” and “contract policies”, to allow this to be factored
28 into an SDI recommendation.
 - 29 • SDI models, in detail, the economic value which each scenario provides to the
30 collective vendors and to each individual vendor. SDI can inform each individual

vendors of its own predicted value, while withholding information about other vendors.

- SDI can seek coalitions of vendors with similar interests to combine their bargaining power, and attempt to compel vendors to enter particular preferred value chain scenarios.

SDI is privy to most or all of the information relating to all of the companies in a supply chain, and can be trusted with the task of creating strategic plans between different companies, potentially making the supply-chain more efficient and benefiting all vendors.

The strategic plan might suggest:

- 1) New business relationships between the most synergistic entities
- 2) Contracts for existing commercial relationships which embody and enforce specific terms of those relationships within the framework of this optimal supply chain system.

SDI might also create a detailed system for accounting for the effect of a contract on the rest of the supply chain (individually and collectively), and predict short and long-term effects of contracts in advance. This information can be used to encourage the desired contractual activities on the part of vendors in a chain. There may also be cases in which the direct/indirect economic impact of a contract upon a vendor results from an additional secondary economic advantage to that vendor besides the trickle down economic effect; for example because of strategic commercial benefits. These factors can also be considered in an economic benefit model.

Given models to compute the comparative economic value of proposed contracts to vendors in a supply chain we can demonstrate a “trickle down” economic value chain in which the success of company A is directly related to the success of company B. This information can be presented to company A and each intervening supplier. SDI can then establish long term contracts between each entity in the chain to insure that each

05699098-10200

1 company receives the benefits of products and services from its other suppliers which are
2 of greater value than without the contract. This trickle-down benefit can be used to entice
3 a company to enter into a recommended contract, for its own benefit and the benefit to
4 company B.

5

6 The benefits which each vendor is predicted to receive in a proposed value chain can be
7 pooled together and used to “bribe” each vendor within the value chain to enter into the
8 preferred contract. While some value chains may actually be willing to provide greater
9 compensation to a vendor, the goal is to link a vendor’s decision to the value of the
10 *complete* value chain, both future and present.

11

12 Of course, vendors may face risks in committing to long-term contracts, particularly
13 when many vendors in the chain must prosper for the value to be realized. SDI might
14 introduce a number of decommitment clauses to allow a vendor to back out of a plan,
15 perhaps reverting to a “plan B” which protects the interests of harmed vendors while
16 retaining as much value to other vendors, i.e. allowing all vendors to contribute to the
17 cost of failed contracts. Alternatively, we might create a *futures market* in which the risk
18 of the future value can be traded in *real options*, within a public market—providing a
19 financial instrument to share risk. In this case, if some measurable form of economic
20 benefit such as sales value (**for e.g., the particular relevant product line**), possibly
21 stock values, profits, etc. falls below a target level the vendor may be compensated for its
22 initial risk, i.e., secure a worse-case outcome.

23

24 It may be important in high-growth markets such as technology, with high risk
25 companies, to allow insurance for contractual non-performance of suppliers.

26

27 SDI can leverage available data to compute the most beneficial economic scenario for all
28 associated vendors collectively, and also to compute optimal scenarios for each
29 individual entity. SDI can disclose the complete picture to each agent, and then allow
30 individual vendors to optimize their local plans in a global plan. To effect a new scenario
31 though, agent agreement is required. Another mechanism could have SDI report the

Each vendor must provide SDI with as much information as possible about the particular conditions under which it will perform new contracts, such as its preference between short term and long term predicted benefits, etc. SDI looks for outcomes that are favorable to individuals and to the overall system. Value can be realized in short-term payments and long-term bonuses.

16 *Additional examples include:*

- 18 • Company A may threaten to file a lawsuit against company B (which could be
19 very damaging to company B). In a similar barter exchange, company B can
20 **barter** its own value (in the form of upside or products/services) as offered to the
21 customers (or vendors of the customers) of company A, targeted to only those
22 customers that reside within the same supply chain. Thus, by limiting the value
23 proposition to only these entities, and not to those which reside within the supply
24 chain of a competitor, the value of the potential upside is worth more because it is
25 partially predicated upon their own direct success. Again, it is also possible to use
26 an indirect strategy, bribing entities or individuals of influence within a given
27 company from which a particular action is desired.
- 28
- 29 • In another novel application there are certain activities among certain individuals
30 which are considered to be mutually and collectively advantageous or
31 disadvantageous to all members of a certain vendor community. We can have all

1 contract all the relevant strategic partners before a funding commitment, to
2 provide more information and reduce risk. Other equity shares could flow from
3 marketing agreements, e.g. a co-branded affiliation of all further
4 products/services. The new company might also guarantee an exclusive contract
5 with another entity. An entrepreneur might also be willing to sell a company to
6 the more established entity, e.g., within a year or two, this can be contracted at the
7 outset. The present framework allows value from long-term strategic alliances to
8 be released.

9
10 Within SDI we can adapt a user's personalized portal browser to favor the vendors
11 providing such value opportunities to the consumer. A vendor could achieve additional
12 market share by complying with requests of another vendor. The value proposition as
13 presented to the user can leverage the personalized browser user interface to prioritize a
14 vendor's offers in the future.

15
16 *Example Commercial Application.*

17
18 An investor has invested in a very high risk venture relating to highly speculative stem-
19 cell research. After years and millions of dollars, the venture finally goes out of business
20 (or becomes acquired on the verge of bankruptcy) the investor loses all of her investment,
21 however terms in the original investment contract (which involved at the time the
22 approval of a variety of medical organizations most likely to benefit from the eventual
23 medical science benefits of the research) now enable the investor to acquire in kind
24 benefits (which may include anything from equity in these various medical organizations
25 to goods/services provided by them). A portion of the value may even include a sizable
26 amount of value provided by the original company and/or indirectly vis-a-vie the
27 remaining medical entities, a substantial amount of value within the barter exchange for
28 favors.

29
30 With time, stem cell medical applications proliferate thus, the "upside benefit" increases
31 in proportion over time. However, if this "up-side benefit" (of the original collection of

1 Thus it is a primary objective within the scope of the present system's formation of
2 common political support strategy groups to ascertain each individuals user's ultimate
3 political objectives based upon the analysis which provides analysis and prediction of the
4 present and likely future effects, both politically and economically from the standpoint of
5 that individual (which may include social, professional and predicted individual investor
6 oriented direct and indirect consequences at a detailed level as it relates to that individual.
7 And secondly, to formulate a global political strategy based upon these objectives which
8 are represented by an SDI agent and which is able to recommend actions and, as such, act
9 as a coalition on behalf of the constituent individuals, and including negotiations for
10 further coalition formation on various portions this global strategy in order to further
11 enhances the collective power wherever common interests can be safely identified and
12 shared to both groups' mutual best interest.

13
14 The effects on commercial industry, even specific businesses in which certain actions are
15 likely to result, may be modeled and presented to the user as well as the effects upon the
16 user in light of the investment portfolio of the user. If a particular political position were
17 to be taken, the system could even recommend-----
18 -----which is most
19 compatible with a particular political position.

20
21 Continuation to Supply Chain Section

22
23 It is also possible to utilize the extensive information in SDI to instruct each political
24 support group's most advantageous supply chain structure in order to be able to
25 recommend certain strategically advantageous cases of tax dollars which facilitate the
26 construction of such supply chains in accordance with the above methods above
27 described. It should be noted that a "political group" may be further subdivided into
28 groups based upon purchase affinities (which themselves suggest different group
29 divisions) which in turn may bribe or boycott commercial entities as needed to achieve
30 their most advantageous strategy as consumers or simply individual users (as economic
31 entities). Each subgroup (via SDI) may further augment their economic leverage with

1 their political leverage (over tax allocation decisions). E.g., if candidate X supports this
 2 particular supply chain strategy and if he is elected then it is most optimally advantageous
 3 to also utilize consumer leverage in supporting transactions with certain vendors and
 4 boycotting others in order to support or complement the politically (tax) funded strategy.
 5

6 As suggested, SDI may be used to facilitate the matching and strategic formation of
 7 coalitions of groups which are most likely to benefit one another on particular political
 8 issues or formation of strategic contractual relationships between commercial entities
 9 (SDI is able to determine if/when certain information is helpful or harmful to disclose to
 10 another particular entity). Each group may exert particular leverage over politicians. I.e.,
 11 not only could the politicians receive feedback as to the political issues and economic
 12 (commercial) alliances which are most advantageous but also the groups could upon the
 13 advice and direction of SDI make their votes conditional upon the politician supporting
 14 certain issues or economic actions. SDI could provide real-time feedback right up until
 15 the election as to the numbers and percentages of users who had supported the candidate
 16 and the candidates response to them. There are other instances in which the leverage
 17 could be used to incentivize commercial entities to support certain political concerns e.g.,
 18 if company X donates percentage of its revenues to the United Negro College Fund group
 19 Z will support the facilitation or an alliance between company and a major manufacturing
 20 firm (e.g., by putting consumer pressure upon those retailers which are the
 21 manufacturers' customers to in turn threaten temporary boycott of the firm unless the
 22 stipulations are met or if they are not consumers to instead offer consumer loyalty (e.g.
 23 through dedicated currency).
 24
 25

26 A variety of applications of the present scheme are conceivable such as local politics such
 27 as commissioners office and board members delegated for purposes of masterplaning and
 28 use applications and allowing or disallowing certain real-estate and highway development
 29 projects. Certain commercial projects, for example, could be of substantial benefit to
 30 certain commercial entities which could provide considerable direct long term economic
 31 benefit to citizens in the local community even directly to the individuals in the

1 communities (who elect those officials) as well as contribute indirectly via the available
2 tax base. Or certain commercial entities could, for example, directly or indirectly provide
3 economic benefit to certain individuals, e.g., if they are employees, investors or
4 commercial entity which is in some way commercially dependent upon that company,
5 e.g., for commercial business if it is a customer or alternatively a supplier and the impact
6 of its resulting success in general (or even local presence) would again provide trickle-
7 down economic benefits to that entity. In accordance with the present application it may
8 even be possible to estimate values for such factors as, for example, the economic impact
9 20 years in the future upon commercial business and industries in the local region (as
10 well as nationally) if local schools are provided with a 10% vs. a 7 ½% share of the local
11 tax base. With these predictive models, the associated prospectively affected commercial
12 entities could, effectively encourage SDI acting on behalf of the local citizens to form a
13 coalition which makes their voting of the relevant elected official(s) contingent upon the
14 increased tax allocation for local schools (the specific preferred users could be predefined
15 as well). SDI acting on behalf of the interests of the prospectively affected commercial
16 entities could bribe the citizens with either (preferably) such benefits as in kind,
17 goods/services which could in turn be leveraged through the supply chain or through a
18 barter exchange system or employment benefits or wage increases (which again could be
19 effectively treated as a barterable commodity subject to the above conditions of avoiding
20 interactions with competitive entities. Such modelling could be extended to many other
21 situations, e.g., allocation of funds to police, fire and EMS services, approving and
22 providing highway improvement again may provide the above benefits etc. providing
23 safety and preserving the health, welfare and life (in addition to educational quality) of a
24 certain number of individuals (each with a certain predictively modeled) local economic
25 impact. The cost versus impact affecting each citizen could be presented to them by SDI
26 accordingly in order to elucidate their decision making processes.

27

28 There may be certain instances in which users at an individual level may be able to provide
29 predicted -----this benefit can be quantified towards commercial entities or
30 even individual users if particular events can be achieved (or particular events prevented).
31 For example, consider a student who is determined to be capable towards achieving

1 substantial real-world success. In one instance the student cannot afford tuition for graduate
2 school. In another instance, the student cannot afford to continue his/her studies due to
3 depleted financial resources. The present system may model and predict likely beneficiaries
4 commercial or even individuals who are likely to benefit and accordingly "insure" certain
5 long-term monetary benefits (if a commercial entities) or even personal or social benefits (if
6 an individual).

7 Internet piracy of copyrighted information is costing authors billions of dollars per year.
8 Technical experts (e.g., within SDI or outside "domain experts" which SDI had identified
9 and presented the problem) identify a technical solution consisting of a ground breaking
10 technology which could greatly reduce total piracy from occurring, however, such solution
11 would require cooperation from all of the Internet server manufacturers. There is, of course,
12 the small risk that in the relatively short-term the technology will partially or completely fail
13 (e.g., copyright pirates find a way around the solution) before a sufficient amount of
14 revenues are realized by the copyright owners to off-set the cost, and justify the expenses to
15 the server manufacturers, of integrating the technology into their services.

16
17 In addition, if copyright owners receive their rightful share of royalties, more Web servers
18 will be sold in general to host such content. (Insert 2 from page 3) SDI can be a facilitator to
19 enabling a beneficial solution to the industry in the following ways:

- 20 1 Present estimated industry-wide revenue losses compared to the industry-wide -----
21 -----the server manufacturers as a group sustain (based upon sales data each one
22 individually and secretly submits to SDI).
- 23 2 Also present the increased sales the industry would sustain if the copy-wright owners
24 received their rightful royalties to the extent which the solution is able to provide.
- 25 3 Provide insurance or create a future market to insure investors who would invest in the
26 commercial opportunity for a share of the anticipated monetary benefits to the industry
27 (such investors could include members of the industry).
- 28 4 Negotiate a deal with investors and/or an insurer (the latter of which could also be SDI
29 representing the industry). A SDI again, of course, may also represent the investors.
30

1 We can allow the community dollar/iamworthit model to extend to physical transactions, for
2 example via smart card technology. A buyer can carry a smart card with community dollars,
3 that are loaded from the user's computer device, or from some third-party operated kiosk.
4 For vendors who offer the integration of community dollars into their discount programs
5 (e.g. may be members of a vendor coalition) the buyer may at any vendor terminal, kiosk or
6 his/her PC upload targeted-vendor promotions such as digital coupons which are tailored to
7 the users profile. This profile is downloaded to the vendor and which consists in part of the
8 recent purchases s/he had recently made at the various vendor stores.

9 In another variation, the community dollars may be coded into a form which is bar code
10 reader enabled and distributed to the buyer electronically or potentially if used in
11 conjunction with a traditional loyalty points program, additionally printed for the buyer at
12 the vendor's physical location (such as point of sale or kiosk), applied in conjunction with
13 purchases at the vendors physical location. At which point typically a new coupon is
14 reprinted containing the updated secure information pertaining to the buyer's community
15 dollar and/or loyalty points account.

16 In another variation, a promotion for a yearly allowance of community dollars could be
17 printed as an advertising offer on a magazine coupon, newspaper insertion or direct mail
18 piece which could contain a unique URL (typically the actual URL for the iamworthit
19 community dollars subscription site with a unique post script as the character string ("code")
20 identifying that particular vendor and/or that promotion) from which the buyer could
21 subscribe to iamworthit, wherein the unique URL acts as an identifier for that particular
22 vendor's promotional piece from which the buyer originally received the offer for his/her
23 own community dollars promotion.

24 Within SDI we could send physical solicitations to buyers, and allow buyers to access
25 promotions pseudonymously. SDI can target a selected audience for each vendor.

26 Example: An iamworthit card in accordance with the pseudonymous payment methods
27 described above, such a card could be a direct extension of SDI into the off line
28 environment. Buyers could use this card as an identifier such that when they travel
29 physically from vendor to vendor, their profile data can be readily identified where data
30 pertaining to their own behavior and policy (depending on their data release potentially part
31 of the vendor's user profile data) is retrieved.

1 If a smart card is used this user profile data may not have to be remotely retrieved but may
 2 be stored on local memory on the card itself along with the client-based pseudonym proxy
 3 server. In one novel variation, a card is done away with completely by virtue of
 4 revolutionary technological breakthroughs in being able to instantly and positively identify
 5 buyers biometrically using iris scanning techniques (which may in a variation be further
 6 combined with facial recognition techniques). Many vendors will wish to utilize user profile
 7 data in order to deliver promotions targeted discounts and promotions (see pending patent
 8 "System for Customized Prices and Promotions")

9 **1.7.1 Location Enhanced SDI System**

10

11 The co-pending application entitled "Location Enhanced Information Architecture" (LEIA)
 12 describes an integrated advertising delivery platform which selectively targets buyer
 13 personalized advertising based upon both the buyer's personal profile and the present
 14 location of the buyer which may suggest appropriate ads from vendors which are local to the
 15 buyer, wherein user identifiers (UID's) which could include any of the above identification
 16 media provide the essential elements for this buyer targeting platform. With LEIA
 17 information providers can utilize location information, in addition to static and dynamic
 18 profiling information. The method customizes the information that is displayed on a private
 19 or public information device to the real audience in the vicinity of the device, instead of a
 20 predicted audience. LEIA collects an extremely detailed and comprehensive information set
 21 about the daily activities of a user, enabling enhancement of the user profile with location
 22 information and temporal activity patterns. The co-pending LEIA patent suggests
 23 appropriate application environments, for example in a smart home, an office, on a mobile
 24 shopping device, and in an automobile. A LEIA-based system stores personal information
 25 on users.

26 We can extend LEIA by incorporation with the Secure Data Interchange system that we
 27 teach in this patent. SDI enables the user to receive the benefits of powerful and well-
 28 directed information, but within a system that respects his/her privacy requirements. The
 29 interchange acts as a secure data warehouse for users and information providers, enabling
 30 information providers to target users without revealing private information to the providers
 31 directly. As described in LEIA, the privacy architectures provided for in LEIA and SDI are

1 all the more critical as components of an SDI enhanced LEIA system architecture given the
2 extreme sensitivity of personal location data which must be securely protected in both real-
3 time and within an asynchronous context.

4 LEIA customizes information that is displayed to an information recipient based on object
5 profiles and physical location of users. Presents the information most relevant to the REAL
6 audience, not a predicted audience per se. Because of LEIA's ability to combine user's
7 preference information with information about their current physical locations and deliver
8 advertising which is both of personal and location contextual relevance to the user, LEIA
9 provides a valuable enhancement to the targeted advertising services provided by
10 iamworthit. For example, instead of advertisers being limited to accessing desired
11 preference attributes of users in the pseudonymous user database exclusively they may
12 additionally access prospective target recipients by present (or anticipated future) location
13 parameters of users connected to mobile and even terrestrial communications networks.

14 In accordance with the iamworthit targeted advertising platform either advertisers select user
15 profile attributes which they are interested in or the system matches their ads automatically
16 to the most relevant users. The former application may be performed either within the
17 context of the pseudonymous user database which the vendor queries or targeting rules
18 which s/he provides which control the advertising and pages which are dynamically
19 generated for each user based upon his/her user profile. Whether the user's location is
20 detected on a mobile network or a terrestrial network including telephone or cable TV,
21 LEIA's location features are able to provide additional useful features regarding users to the
22 vendor's rules interface or to the pseudonymous user database controlling dynamic page
23 generation at his/her site.

24 For example, at a bookstore, we can recommend isles and particular books; at a
25 supermarket, can play music preferences; smart-radio, play appropriate channels in a cab
26 based on target object profiles (as meta-data). As suggested in issued patent "System for
27 Broadcast of and access to Video and Other Data Using Customer Profiles" appearance of
28 relevant selections can be continuously scanned for, dynamically selected and presented to
29 the buyer in the form of "Virtual radio station". Such a system can also be linked to a
30 service for making an instant purchase, or linked to a database (in conjunction with LEIA) to

1 recommend where a buyer should physically go to make a purchase. For example, music
2 selections that the buyer is presently listening to may be ordered.

3 Also, it is possible to provide advertising which is targeted to a buyer by automatically
4 recognizing pre-existing commercials and replacing them with targeted counterparts. This
5 can be done through the identification of previously played commercials for example,
6 commercials that have been manually identified and classified. Upon recognition, targeted
7 commercials (including those which are targeted by buyer location in accordance with
8 LEIA) may be inserted into these spots, and delivered and/or pre-cached through cellular,
9 satellite or radio communications.

10 At a public phone we can identify a buyer with his/her calling card, and deliver targeted
11 advertisements, via the public telephone readout or delivering the targeted ads as audio
12 messages in which server software at the phone switch (an ISP level proxy) recognizes
13 if/when the buyer is put on hold and delivers audio and/or audio/video advertising to the
14 buyer accordingly.

15 Targeted discounts and advertisements can be delivered at kiosks, for example using a credit
16 card/smart card/other ID method (e.g. biometric...) Similarly, we can use credit cards to
17 deliver targeted print advertisements on the backs of purchase receipts, e.g., supermarkets or
18 fueling pumps or, alternatively, on a sheet dedicated for a advertising purposes conjunction
19 with public copiers or printers or in another variation, on the cover sheet of incoming faxes
20 which are sent to the buyer's fax machine or in which the buyer is otherwise identified
21 automatically from the recipient's name field on the cover sheet.

22 One application includes "Smart Home Intelligence", where methods are disclosed by which
23 users' real-time behavior may be profiled through their movement throughout their home,
24 and specific interactions with the various network enabled appliances throughout the home.
25 Other inputs may include the user's speech patterns (using voice recognition in combination
26 with text analysis). It could for example, note the user's speech content patterns in real-
27 time. Such information provides invaluable clues as to the user's present activities, mood
28 and interest state and may be processed by the presently described algorithms tuned with
29 location/time features typically using the assistance of human data analyst to identify the key
30 features and correlations. (This information may also provide enhanced information
31 pertinent to the user's general, static preferences as well).

0969098-102700

1 Other extensions of this scheme are also considered e.g. within the context of the user's
2 office, or automobile and pedestrian activities. This application may thus extended the
3 usefulness of the iamworthit model to advertisers in being able to target users through the
4 presently anticipated on-line media as well as networked appliances and in either case, based
5 upon the relevant context of users' present activities and behavior (and from this potentially
6 their inferred moods or mental states) within their homes and elsewhere. Additionally
7 (perhaps most importantly), LEIA provides a highly beneficial value to vendors whose
8 customers purchase primarily from the vendor's bricks and mortar store presence. SDI can
9 identify multiple vendors which share a common physical location. Additionally, LEIA can
10 also identify the physical location of buyers on mobile and terrestrial networks and suggest
11 the sharing of off-line prospects for these off-line retailers on the basis of not only interest
12 but physical location.

13

14 **1.7.2 Digital Set Top Box Methods**

15

16 Similarly, we can deliver targeted advertising and other information through cable TV
17 systems, as described in the issued parent patent application entitled "System and Method
18 for Scheduling Broadcast of and Access to Video Programs and Other Data Using Customer
19 Profiles" US Patent # 5,758,257, and co-pending application entitled "Broadcast & System
20 for reduced memory terminals broadly address the use of cable systems as an interactive
21 medium (in a bi-directional network architecture) for purposes of delivering targeted
22 advertising targeted advertising and other information to the consumer based on user
23 profiles".

24 In this system customer behavioral data is collected at the digital set top and the upstream
25 channel enables these profiles to be processed at the lead end server. These detailed profiles
26 may then be subsequently transmitted down and stored at the level of the individual set top.
27 The cable environment is a two way interactive medium. The bandwidth allocation is
28 inherently asymmetric. Separate channels can push parallel adverts, which are selected at the
29 set-top-box according to a buyer's profile. Each channel can have associated meta-data to
30 allow matching at the set-top-box. As an alternative variation, full motion advertisements
31 may be down loaded in the form of applets to the digital set top box and displayed to the

1 buyer in similar fashion as described above. Or full-size commercials or infomercials could
2 even be downloaded and inserted in place of existing commercials on the video stream.
3 Digital tags used for queuing for ad insertion technology enable this capability. The
4 preferred commercial implementation of the system is used within the context iamworthit
5 (the buyer infomediary service in which the viewer receives value for advertising received).
6 This method is preferred (in as much as in the following alternative, the cable operator does
7 not need to be in the loop). With that said, in an alternative embodiment, however, requiring
8 the cooperation of the cable operator various commercials could be broadcast in parallel
9 during commercial breaks. The user profile at the set-top is deployed to select the most
10 appropriate commercial accordingly.

11 User profiles regarding viewing behavior can be collected with relatively minor up-stream
12 transmission, e.g., to infer whether a buyer is viewing a program the set-top may transmit
13 the viewers current viewing selection, e.g., two minutes after the beginning of each half hour
14 and two minutes before the end of the half hour (and possibly at additional intervals during
15 the viewing segment as well). In the point-to-point access and delivery of personal
16 information as well as the passive monitoring of viewing behavior (for profile generation at
17 the head-end and subsequent user profile delivery for the set-top) because of strict buyer
18 privacy regulations in the cable/satellite industries the use of the pseudonym proxy server
19 architecture (as described in the parent case and integrated into the above referenced case) is
20 extremely important.

21 Once interactive television is enabled by access to very large amounts of bandwidth, it will
22 be very feasible from a bandwidth utilization standpoint to transmit more interactive
23 content. Predictive caching is still extremely advantageous in as much as advertising (unlike
24 interactive video real-time on demand video content) tends to be controlled by targeting
25 rules of the advertiser rather than the user who receives value in exchange. Of course, radio
26 or music content may also be delivered in conjunction with preloaded audio commercials.
27 And any of the above content is deliverable over the same networks to other appliances such
28 as the P.C.

29 **1.7.3 Optical-Based System Extension**

30 Iamworthit in a future version of the system could incorporate a novel optically based
31 medium for delivery of the advertising. The protocol would enable the delivery of user

09699098 "102700"
1 targeted advertising in a non-electronic environment directly to the user from a static source
2 in visible contact of the user. The system incorporates the following aspects: The user
3 wears glasses or contact lenses which contain a Polaroid film in which the optical medium is
4 polarized for only certain very narrow and specific wavelengths. Advertisements (or other
5 messages) may be presented from signboards which display from the same surface ads
6 appropriate for every different user profile segment for which a unique message is
7 appropriate. The Polaroid film for each user is unique to the user segment to which they
8 belong. Each user segment's optical film filters out all of the particular wavelengths of all
9 other messages except that which is appropriate to the user segment to which that user
10 belongs. Alternatively, the current application may also be relevant to virtual tags as users
11 may provide instead optical messages specifically targeted by a user to other users
12 characterized by their profile features which establishes the entire criteria for the user
13 segment they fall into for which an appropriate message may be targeted.

14 **2. Interactive Data Mining Applications**

15 In this section we describe interactive data mining applications, which involve
16 dynamic two- or multi-way communication between agents. For example, within the
17 context of a vendor-consumer interaction, the key difference is that information flow
18 is not one-directional. An important example is time-of-purchase, in which the
19 consumer's agent initiates a request for a response (with a counter bid) from a number
20 of vendors. The vendors can access profile information about that consumer, and then
21 make personalized responses. In general, this bid-response processes can iterate, with
22 vendors/consumers exchanging information over multiple rounds. The exchange of
23 information may occur within SDI in initial stages, with the system acting as a trusted
24 intermediary until the agents are ready for an introduction.

25
26 We describe time-of-purchase competition, and then move on to general match-
27 making applications, which allow interested parties to exchange information by
28 mutual consent (as indicated in SDI rules), with initial information exchange
29 autonomous. There are a number of interesting match-making applications, for
30 example within a system for sensitive negotiation, introductions. The system of SDI
31 can play the role of a trusted intermediary, so that only suitable introductions are

made. A vendor can place a request for a certain type of meeting without its competitors knowing that it has made such a request. Finally, we describe a “real-time experts market”, in which experts can respond for payment to questions placed by other agents. The system of SDI allows a useful matching of agents and experts, and also allows data mining to check on the feedback a particular expert has received from previous clients. We also suggest an application of interactive SDI-based data-mining to resale markets, and to a decentralized transportation management system.

2.1 Time of Purchase Competition

One application of the system of Secure Data Interchange is in “time of purchase” competition, where by a networked system of vendors can subscribe and receive the opportunity to place counteroffers to users that are about to make a purchase. This application is enabled within SDI because the user agent can remain anonymous while receiving counteroffers, but still use the system of SDI to provide profile information that can allow vendors to make attractive personalized offers. The client-side SDI proxy notifies a central SDI “time of purchase” server, which can:

- Determine appropriate vendors to prompt with an opportunity to make a counteroffer.
- Create a one-time anonymous identifier for the user, linked to an appropriate set of profile information, and allow vendors to execute queries under that profile.
- Collect counteroffers, and pass them to the user’s client machine.

We also describe how to generate coalitions dynamically, based on the privileged position of the time-of-purchase server in the marketplace. The time-of-purchase server can transparently bundle purchases from disparate end-users, and negotiate quantity discounts with vendors. Buyer Coalition formation can significantly improve negotiating leverage as the collective buying power of numerous individuals using the time-of-purchase competition feature, and be made completely transparent to users.

Time-of-purchase allows a user to elicit dynamic market competition between vendors. For example, when a user browses a book at Amazon.com if she/he chooses to activate the time-of-purchase competition feature from the browser, a notification will be sent-out in real-time to all competing book stores which also offer that product section. The notification contains

1 the product code and price quoted to that user at Amazon. The electronic agent at the
2 vendor's site has an opportunity to counter Amazon's offer to the user (typically on the basis
3 of price, however, other parameters may be used as well). The user may then select the
4 offer which is most competitive or attractive, including the original offer.

5 Integration into a standard shop-bot interface e.g. with comparative features of the products
6 associated with the various offers across a variety of product criteria, or at the least price can
7 be used to provide the user within simple recommendations.

8 Another extension is to propose a special interest-bearing bank account which we set-up for
9 the user which is tied into a wallet or debit/credit card for off-line transactions. We can
10 allow a user agent to bundle its own purchases over time, and use the system of SDI to
11 guarantee future payment to on-line vendors in return to a good current offer. This is novel,
12 because it is like participating in a "discount scheme" without purchasing the right to future
13 discounts up front. A user can still invest money that is earmarked for future purchases with
14 a vendor.

15 **2.1.2 Description**

16
17 Figure 17 illustrates the overall architecture of this on-line B2C e-commerce application of
18 the methods of SDI. The agent represents a user with an SDI-enabled client machine, the
19 vendor a server computer, also SDI-enabled. Agents interact with vendors both directly and
20 via the central SDI server, and other components. We illustrate the time-of-purchase server,
21 which receives requests from user's client machines as purchases are about to be made, and
22 forwards those requests via the SDI central data warehouse to vendors.

23 Client-side proxy agents monitor the interaction between a user agent a vendor, and can be
24 configured to automatically notify other vendors when a purchase is about to take place. The
25 key problem that we address to implement a system for time-of-purchase competition is
26 how to identify the type of purchase that a user is about to make, automatically. This makes
27 the operation of time-of-purchase seamless to a user, the system simply requires that the
28 user waits for a period of time while competitive bids are collected from other vendors.

29 We propose two variations:

- 30 a) The vendor with which the user is about to make a purchase subscribes to SDI
- 31 and provides information to the client-side proxy about the user's interaction.

1 that encodes relevant information about the type of product the user wants to
2 purchase.

3 3. User clicks on ‘time of purchase competition option’ on its local SDI-enabled
4 client.

4. Client sends message to the SDI time-of-purchase competition server with (User profile information/identifier, one-time time-of-purchase ID, product/service classification code)

5. Time-of-purchase competition server looks up the product code, and sends the user profile identifiers and a time-of-purchase ID to relevant vendors.

6. Vendors can construct competitive offers based on profile information accessed in the central SDI data warehouse, and information about the existing offer.

7. Vendors send the counteroffers to the time-of-purchase server, which forwards the offers to the user's client-side SDI proxy.

14 8. The user is presented with a set of alternatives, and can make a final purchase
15 decision.

16 9. The final purchase is completed with the appropriate vendor.

We allow other vendors to access profile information about a user based on the profile information that the user provides to time-of-purchase, i.e. based on the profile which it wishes to use for the purposes of having appropriate counteroffers constructed. At this stage the client-side SDI proxy agent might also send additional profile information to be stored within the SDI data warehouse.

23 It is useful for vendors to access profile information about a user (anonymously or
24 pseudonymously, as defined by the user's proxy agent) because products and services can
25 be configurable goods or services, and a vendor can compete on more than price alone, but
26 also in other dimensions.

27 The one-time time-of-purchase ID is constructed client-side in step (4) to act as a temporary
28 identifier for the purposes of soliciting competitive responses from vendors. Vendors can be
29 prevented from making direct offers to users, all counteroffers must be forwarded though
30 the time-of-purchase server.

1 4). B2B Commerce – Numerous commercial applications including standard Web-
2 based environments for B2B e-commerce including buyer-side and seller-side e-
3 commerce applications, vertical B2B portals and trading floors.

4 5). Job Positions – Employers posting job positions may be notified by the service of
5 competing job offers which employment candidates are browsing (and ultimately
6 accept if this information is indicated by the candidate). Resumes and, if desired,
7 other profile information is provided to all competitors. Depending upon the profile
8 of the candidate and/or competitive employment opportunities which a particular
9 candidate is observing (or has recently observed), employers may wish to customize
10 offers dynamically on an individual basis.

11 **2.1.4 Integration into a Shopbot Interface**

12
13 We also suggest the integration of time-of-purchase competition into a standard shop-bot
14 interface e.g. with comparative features of the products associated with the various offers
15 across a variety of product criteria, or at the least price can be used to provide the user
16 within simple recommendations. With this, even without a time-of-purchase offer the
17 user can at least compare its offer with fixed-price deals in the wider market place. We
18 might also provide a user with historical information about previous (low) price offers,
19 within a Yahoo-like portal for e-commerce; to provide a user within information about a
20 reasonable price for each individual purchasable; i.e. the very lowest price that each
21 given item had been offered to a customer previously using time of purchase competition.
22 This site could also act as a stand alone proxy server (like anonymizer) which inserts this
23 information by overlaying it on Web pages as the user browses the web.

24
25 The time of purchase competition feature is designed to be accessible to a user within the
26 context of or most any information access mode relating to browsing on the Internet e.g.
27 Web or portal browsing, receiving email or “push” content, submitting queries (e.g.
28 specific names, categories and/or desired features (or combinations of the above) of
29 products which the user identified as being of present interest.

30

1 During any of the above information access modes when a buyer specifies *time-of-purchase*
2 *competition* at her SDI client proxy, SDI can automatically provide competitors with
3 information about a buyer's product or service requirements, and a buyer's profile, as well
4 as on particular vendor specific offers the buyer is observing or contemplating before a
5 buyer makes a purchase. This can facilitate competition between vendors, and can lead to
6 better prices and offers for buyers. The system of time-of-purchase competition can also
7 help to reduce the costs of entry into a market, because name-recognition becomes less
8 important. New vendors can simply register with *time-of-purchase referral system*, and
9 cherry pick a small set of purchases.

10 This is a next-generation e-commerce service. Current shop-bots, for example "Junglee" at
11 Amazon.com and www.shophtheweb.amazon.com provide a *static* comparison shopping
12 service. A buyer can specify a product, and receive price information about the product from
13 different suppliers. There is no dynamic competition between vendors on price or features.
14 The buyer driven service for flights offered by www.priceline.com is more dynamic, in that
15 a seller is found to match the price that a buyer bids, but does not necessarily promote
16 competition between sellers for a user's purchase. In fact the sellers can make excess profits
17 from the pricing errors made by buyers, because once a buyer states his/her reservation
18 price, that is the price he/she must pay. Surplus goes to the market operator, e.g. priceline,
19 but not to the user. The system is also different to our proposed system in that:

- 20 a) Time-of-purchase competition is disintermediated and decentralized,
21 implemented dynamically over a virtual network of competing vendors.
- 22 b) Time-of-purchase also allows a vendor to access anonymous profile information
23 about a buyer, to enable a careful tailoring of an offer to the preferences of a user.

24
25 We can use profile information, and historical transaction information for similar
26 transactions, together with the customer price/promotion algorithm disclosed in co-pending
27 patent "System for the Automatic Determination of Customized Prices and Promotions" to
28 negotiate on a deal with a vendor that will optimize the value to the buyer. Profiling of
29 vendors, and buyer transactions, can allow buyers to avoid making bids that are too high and
30 losing value (airlines in priceline.com can profit from inaccurate buyer bids).

1 as disclosed in co-pending patent application "System for Automatic Determination of
2 Customized Prices and Promotions" and U.S. Patent #5,754,939, "System for Generation
3 of User profiles for a System for Customized Electronic Identification of Desirable Objects"
4 as such techniques may be deployed by the vendor directly or via the Secure Data
5 Interchange representing the interests of the vendors.

6 *Choosing an Offer to Make*

7 User profile information may include a temporal profile of the buyer's present activities,
8 including search terms, recent page navigations, what pages is the buyer observing presently
9 (and the profile of this page) or even his/her present physical location as well as the general
10 user profile. Any portion of the above particularly the latter two may of course be withheld
11 from the vendor). Or the buyer may be more generous in providing this information about
12 the various static or dynamic features.

13 Vendors can target buyers on the basis of their preferences and interests, and also within the
14 temporal context of when they are most likely to be receptive to offers (thus providing also a
15 benefit/service to the buyer as well as enabling the competing vendors to increase the price
16 point at which their competition begins, knowing that the buyer is in an optimally receptive
17 condition to accept offers for that item). For example, a user might release to a vendor the
18 current context of a user's purchase, such as the click stream over the past 5 minutes.

19 In the preferred implementation, vendors are also provided with a (client or web-based)
20 rules interface which enables the vendors to input pre-stated rules with which the system
21 may solicit and respond to competitive offers automatically. In this way data mining may
22 be performed in order for the vendor to determine what a user (or all users) by attribute, tend
23 to best respond to by product, product feature, features or services of the vendor, price, etc.
24 If pre-stated rules are used to automatically respond to a notification with a competitive
25 offer, the nature and degree of discount is typically determined in accordance with the
26 nature and degree of the original or previous offer and/or the user profile as disclosed by the
27 client-level proxy/server to that vendor. In lieu of manually entered rules, co-pending patent
28 application entitled "System for the Automatic Determination of Customized Prices and
29 Promotions" another similar algorithmic methodology may be used as an aid by the vendor
30 in order to automatically determine a competitive offer (or subsequent responses thereto) as
31 well as an aid to the vendor in selecting optimal rules.

1 SDI can also present a buyer with previous time of purchase vendor competition
2 negotiations e.g. offers made in the past by vendors for similar products as those provided
3 by that prospective advertiser (even from the same vendor) which are at least as competitive
4 as offers made to that user or "similar" users. Because of the competitive nature of this
5 environment, vendors will wish to leverage all of this information and using their own
6 pricing algorithms and data mining tools, craft optimal offers through the pseudonymous
7 buyer database.

8 If the buyer so desires, the client-level proxy can also automatically notify these vendors
9 if/when a particular offer is about to be accepted by the buyer. For example, a time delay
10 response in the client-level proxy actually processing the order requests could allow vendors
11 a final opportunity to present another competitive offer to the buyer. In another less optimal
12 variation, vendors are notified only upon the buyer agreeing to accept an initial offer
13 received. A secure function within the buyers client residing within the Secure Sockets
14 Layer (SSL) is able to authenticate to the vendor's receiving the notification that the offer is
15 true and untampered by the buyer through the transmission of a secure encrypted message.

16 As an additional service to buyers the SDI-level proxy server can perform analysis on the
17 offers that a buyer receives, through comparison with offers that have been received by
18 other buyers with the best offer that has been received by any buyer for the same product,
19 and with the typical offer received by a buyer with a similar profile to the buyer. This can be
20 useful to a buyer because it will allow the buyer to reject all offers if they are non-
21 competitive. The SDI-level proxy could also automatically identify for buyers the profile
22 attributes that promote good offers, and the profile attributes that promote bad offers, as an
23 informational service to enable buyers to gain better offers in the future, either through
24 revealing certain information or changing behavior to attain favorable profiles.

25 If iamworthit detects that discrepancies in the pricing for a given vendor's item occurs in
26 correlation with certain identifiable user profile information (not arbitrarily), it will in the
27 future initially disclose that data of any other which is determined from past buyer
28 interactions to be advantageous for the buyer. Should the vendor exhibit price or offer
29 discrimination (i.e. by having provided more favorable items to other previous buyers, the
30 buyer may demand the lowest price. If the vendor ignores the request and the user has not
31 received (or does not wish to accept) another vendor's offer for a lower price and the

1 iamworthit infomediary service may apply pressure upon the vendor such as creating a
2 buyer coalition and/or making a believable threat (as described below) which has the ability
3 to deny future business to that vendor. If the vendor is able to present an exceptionally
4 attractive offer the buyer may instead wish to stipulate that the offer is valid only if the time
5 of purchase competition feature is not used by the user. iamworthit's statistics based price
6 prediction scheme can then predict if this decision would serve the user's best interests or
7 not based upon the user's purchasing objectives.

8 Furthermore, we can allow vendors to offer payment to a client in return for displaying an
9 offer to the buyer, and vendors can also bid for space on the buyer's web portal which is
10 often represented as a profile associated with a pseudonym in conjunction with a description
11 of the ad space. The purchasing decisions of the buyer may be performed by an electronic
12 representative of the buyer's wishes (as "buyer agent") implementing the techniques of
13 pricing/promotion selection algorithms completely autonomously on behalf of the buyer.
14 However, the best offer can only be presented to a buyer to the extent that the SDI client
15 level software understands a buyer's model of "value", and can make appropriate tradeoffs
16 between product features and price (as implicitly inferred by the system through the above
17 suggested techniques or explicitly stated by the buyer in advance). Nonetheless, this is a
18 hard problem, and we expect that the buyer will often need to make a final product choice
19 decision which could be a default in which the system would defer judgment to the buyer if
20 it's statistical confidence as to the buyer wishes below a certain threshold.

21 The collaborative filtering techniques described in pending patent "System for Automatic
22 Determination of Customized Prices and Promotions", can allow a buyer's client-level
23 proxy server, termed the buyer agent in this section, to automatically analyze offers. The
24 system can also be used to send initial offers to vendors, on the basis of historical
25 information about the transactions that have been performed between other buyers (which
26 include the benefit of his/her complete user profile data) and the vendor. Offers can (of
27 course) be sent to a vendor and its competitors. Finally, after offers that are received from
28 vendors are pre-screened, they can be automatically ranked for value—using a combined
29 quality and price metric (again judged within a collaborative filtering framework). Buyer
30 feedback e.g. average ratings and annotations from previous buyers may form an additional
31 criteria. The goal is to leverage the database of other offers that have been accepted by

buyers in the past, and form a model of vendors, to determine whether or not a buyer has received good offers (i.e. we can exchange information within the system of Secure Data Interchange, and making more information available increases the efficiency of the market). Offers can be filtered and presented to a buyer in rank order.

2.1.5 Time-of-Purchase with Non-cooperative vendors

We suggest a simple technique to allow a client-side SDI proxy to automatically detect that a purchase is about to be authorized, and the details of the purchase. The model is that SDI, or a third-party SDI client, provides a banking service for a user. The advantage of such a service is that in making a purchase, we can require that the vendor provides information to the bank about the purchase that is about to be made, as a condition for validating payment.

The system works as follows:

1. User interacts with vendor, and decides to make a purchase.
2. User provides payment method, for example an SDI bank account number.
3. The SDI bank account server demands information about the product which is about to be purchased before authorizing payment for the service.
4. The vendor provides the SDI bank with information about the purchase.
5. The SDI time-of-purchase system can now step into the transaction, and solicit competitive offers from other vendors.
6. The user is presented with the opportunity to select an alternative offer.
7. The SDI proxy-agent reports the user's final purchase decision to the appropriate vendor, and denies purchase to the initial vendor if the user accepts an alternative offer.
8. The final transaction is executed.

2.1.6 Automatic Buyer Coalition Generation

The time of purchase competition system can also be used to allow the automatic formation of user coalitions. Coalitions can be generated dynamically based on the privileged position of the time-of-purchase server in the marketplace. The time-of-purchase server can transparently bundle purchases from disparate end-users, and negotiate quantity discounts

1 with vendors. Buyer coalition formation can significantly improve negotiating leverage as
2 the collective buying power of numerous individuals using the time-of-purchase competition
3 feature, and be made completely transparent to users. The ability to offer a single vendor a
4 number of sales allows significant negotiation leverage, and can be used to get special
5 discounts.

6 Although the coalition formation problem is very difficult in general, because individuals
7 may not have the same goals and the computation of an optimal outcome with self-
8 interested agents is hard (for example, how should the profits from a coalition be divided,
9 how can we assess the value of each agent for the coalition, etc.)

10 The time-of-purchase server has information about the current and pending purchase
11 requests from users in the system, and can make one-shot coalition formation decisions,
12 with profits divided equally between agents. Agents provide a one-time bid for a service, as
13 the offer that they are about to accept from a vendor. This is a guaranteed offer, as the
14 purchase is about to be made. Therefore, it is a simple matter to aggregate bids from a
15 number of agents as a bundle bid for a set of items, and negotiate a good price from a single
16 vendor. This solution to coalition formation is simple, because it is not negotiated over many
17 steps between agents, but performed one-time with agents tied into an ad-hoc coalition.

18 For example, suppose that at any time, in the period of 10 seconds, 100 SDI agents are about
19 to purchase compact discs, varying in price between \$5.99 and \$15.99. The time-of-
20 purchase competition agent has this information, and can perform the following steps:

- 21 1. Collect time-of-purchase requests over a short period of time (e.g. 10 seconds).
- 22 2. Identify purchases with a common type (e.g. compact discs).
- 23 3. Aggregate the sum value of agent's buy prices for the purchases (e.g. sum the
24 price that each agent is about to pay for a compact disk).
- 25 4. Choose a discount on the TOTAL price, e.g. 10%. Make an aggregated buy offer
26 to appropriate vendors on behalf of all clients (e.g. buy all 100 compact disks for
27 a total 10% discount).
- 28 5. If successful, divide the final discount over all clients, e.g. subtract 10% from
29 every CD purchase. Perhaps take a share of the surplus as profit for the brokering
30 role performed.

31

1 The time-of-competition server can also take a position in the market, identifying bundled
2 purchases over time, essentially aggregating demand and making offers of a guaranteed
3 stream of purchases over a period of time based on historical information, again for a
4 discount that can be seamlessly passed onto future purchasers. If the server is successful in
5 taking a position in the market for some good that its user base seem to be interested in
6 purchasing, then in addition to returning offers from vendors in the system, the time-of-
7 purchase system can provide good prices to its clients.

8 Buyers might also form coalitions on the basis of automatically detected synergies between
9 their requests and predicted similarities between them. This can give buyers more leverage
10 in negotiation with a vendor. In the automatic evaluation of how “good” an offer is (how
11 competitively it matches the market demand for that coalition’s needs) in iamworthit’s
12 historical user information, it must also utilize information about the effect that size of the
13 coalition has upon the market’s responsiveness. Within the context of the present
14 framework, the simplest version of the system involves creating the coalition around a
15 particular item and, however, it also conceivable that it could be formed at the vendor level
16 in which the coalition members are interested in any item(s) provided by the vendor and
17 vendor competition is elicited by the coalition across all products which its members
18 demonstrate interest in and which are provided by the competitive vendors. Given a
19 sufficiently large subscriber base, buyer coalitions may be automatically created in real-time
20 and in ad-hoc fashion (i.e. from the iamworthit buyers who are interested in a particular type
21 of item at that particular time).

22 Buyer coalitions typically are able to exert significantly greater pressure upon vendors the
23 larger they are (this fact applies as well to multiple company b to b buyer coalitions). As
24 such iamworthit may also notify other members which in accordance with their user profiles
25 are likely to have an interest in a particular item (or items) in which a coalition of members
26 have demonstrated interest in order to increase the size of the coalition. Similarly, the
27 aggregate user profile of all members of the coalition may be used in conjunction with
28 collaborative filtering in order to recommend those items which collectively best match the
29 preferences of the coalition collectively and thus are able to provide the best deal for the
30 most items possible to that particular coalition. Thus the coalition model is extended to
31 placing competitive pressure upon vendors across multiple items. iamworthit must also

1 be provided on the basis of the profile information directly associated with the pseudonym
2 of a buyer. Certain portions of the user profile data that is unavailable for direct collection
3 by the vendor (such as information that is collected on other sites including, in particular,
4 competitive vendor sites) may reveal important information which enables the vendor to
5 better target that buyer. As such the secure data interchange representing the collective
6 buyers may aggregate, analyze and sell this data to the vendor so long as the release of such
7 information does not negatively affect the predicted pricing or value levels for that buyer as
8 performed by the above type of pricing algorithm.

9 An interesting variation of this example involves the situation in which multiple entities
10 represented by SDI users may be negotiating with the same other entity (or individual). It
11 may be for the same purchasable. Alternatively, it may involve different purchasables or the
12 represented group may even be a combination of users and sellers who happen to be
13 presently dealing with the same entity. Accordingly it is important for SDI to achieve an
14 understanding as to the individual negotiating parameters as well as an assessment of a
15 market demand model which characterizes the needs and objectives of that entity with
16 regards to the particular prospective transactions being negotiated which SDI is mediating
17 on behalf of each party which is captured through the main SDI server. As such it is often
18 possible to thus attempt to predict the minimal acceptable terms of an offer which would
19 meet those conditions. The negotiating leverage from all parties is based upon terms which
20 affect multiple (or perhaps all) of the deals collectively. SDI thus has the power to
21 significantly strengthen the negotiating leverage based upon a market demand model which
22 characterizes that entity thus benefiting the entrusted parties on the whole. The statistical
23 data used for market demand models from the user-centric or vendor-centric SDI is
24 collected about the other parties which the parties are negotiating with. One useful source
25 for estimating this predicted market demand curve which they represent. For example it is
26 possible for end-users as a coalition of users to be automatically formed for this purpose.

27 A user-centric SDI is then assigned to the coalition in order to (a) ascertain from user profile
28 information (in addition to subsequent active query or recommended offer feed
29 back/approval information about terms which the coalition would likely wish to receive
30 from the vendor (subject to negotiation) (b) ascertain from vendor behavior relating to past

1 users and coalitions terms which would be reasonable to expect from the vendor in order to
2 anticipate reasonable offers/counter offers to present to the vendor.

3 The user-centric SDI may even detect from the user profiles the possibility (followed-up by
4 a confirmation query to the users) that the users may be willing to commit in advance for a
5 specified period of time to certain customer performance criteria e.g. in exchange for not
6 only guaranteed discounts but also an equity stake in the company which may increase over
7 time and where continued ownership of which is contingent upon its sustained user
8 performance as agreed upon in advance for the specified period of time. Assuming
9 sufficient willingness, the coalition as directed by the user centric SDI will submit the
10 recommended offer to the vendor along with its competitors using time of purchase
11 competition. The coalition may even incorporate a believable threat (upon approval of its
12 members) (e.g., that it will not deal with the vendor in the future if the terms of its request
13 (such as long-term discount privileges or conversely promise to give the vendor certain
14 competitive advantages in future user coalition negotiations which may include future SDI
15 coalitions.

16 SDI's predictive tool suite can be used to predict the most likely prospects for a given
17 coalition, i.e., matching users who have explicitly indicated an interest or have engaged in
18 negotiations or past dealings with a given vendor. As indicated the recommended users may
19 be targeted either with the same items or completely different items, the latter case in which
20 SDI will try to negotiate a lower price based upon the aggregate sales form the coalition, not
21 the volume of sales for one particular item per se, as in the first instance. In the present
22 application, however, because negotiating pressure is being applied by a collection of
23 entities upon (typically) one entity (or possibly alliance of entities), an SDI (which may
24 typically be spawned from the main SDI service) representing the coalition is created in ad-
25 hoc fashion. The other entity may accordingly wish to acquire its own SDI to also represent
26 its commercial interests. Each SDI may represent multiple parties. No SDI must of course
27 ever represent an entity which is negotiating with another entity which it represents.

28 Coalitions are typically arranged automatically by SDI based upon observed synergies in
29 user and seller activities which SDI observes. In fact the predictive capabilities of SDI
30 could be used to model (1) the anticipated number of cooperating entities which are
31 predicted to join the coalition and (2) from that information, the predicted terms and price

1 coalition's business which is typically a long-term commitment. In conjunction the
2 selection of vendors which the coalition approaches is based upon the willingness of other
3 competitive vendors to conform to the wishes of the coalition with regards to the vendors'
4 selection of suppliers typically. Several multivendor (supplier) relationships currently exist.
5 And the objective of iamworthit is to identify suppliers which are well qualified to satisfy
6 the needs of the vendors which are candidates for their businesses and which similarly
7 provide significant growth potential, if these vendors switch over to these suppliers. If the
8 coalition's offer is successful, it will invest in the supplier(s), prior to its public
9 announcement of having obtained the vendor(s) as a customer(s). If the supplier is in
10 substantial need of investment (e.g. a private placement), it is even conceivable that the
11 coalition may be compensated by the supplier in the form of monetary compensation or
12 preferably free or discounted product from the vendor for which they provide their
13 product/services. This may further assure user loyalty from the coalition and is typically
14 limited to an initial period needed to entice the coalition members that the investment in the
15 supplier is worth while in light of the rather high risk in such an early stage company.
16 Other more elaborate applications are also conceivable, for example, more general requests
17 for knowledge and influence over particular executive decisions which are not highly
18 confidential which may provide benefits to the members of the coalition e.g. in addition
19 other types of partnering decisions with other vendors also potentially to strategic product
20 development directions and revisions which provide the coalition optimal personal
21 satisfaction and value for their purchasing commitments. It is even conceivable that through
22 the aid of vendor credentials (for identification purposes) user coalitions could bargain with
23 vendors using "restricted currency" which stipulates its use must be for particular uses. (or
24 its value for any other users is diminished) (such as vendors in which the coalition has an
25 investment state).

26 In a variation, these restrictions could "encourage" subsequent transacting parties of the
27 currency (e.g. the employees of the company) to buy from particular vendors which the
28 coalition support (e.g. in conjunction with or equity stake), perhaps thus encouraging the
29 employees join the coalition as well. Data mining (performed by SDI) will also indicate
30 existing vendor relationship opportunities for investment between/within (respectively)
31 different vendors as well as the flow of currency through sequences of transactions in which

1 these complex vendor synergies exist even indirectly via one (or multiple) removed vendor-
2 vendor and user-vendor relationships.

3 *2.1.8.2 Coalitions Created For Purposes Of Eliciting Change On The Part Of Vendors*
4

5 In addition to applying market pressures, one other example in which entities may be
6 automatically introduced into coalitions using SDI may involve SDI automatically
7 identifying certain situations or complaints from users or entities in which there may be
8 common ground for these parties to collectively apply political or legal pressure upon
9 another entity from which the problem arises, or in one example to apply legal tactics for
10 purposes of reclaiming damages from the infringing party. For example, members of SDI
11 could be asked to disclose particular instances of these types of legal infractions even if they
12 are relatively minor (e.g. simply a short fall in acceptable quality of misrepresentations
13 through misleading advertising etc. may be relatively minor inconveniences and/or a site
14 may be provided as well for non-SDI members (e.g. www.class.action.com).

15 A similar application tailored specifically to the Federal Government could also be provided
16 (e.g., www.governmentindustries.com. In one example, SDI's ability to provide annotations
17 excluding complaints and ratings pertaining to particular organizations and individuals is
18 complemented by the current scheme by introducing the ability of users to apply pressure to
19 the organization or individual as a coalition in order to more effectively motivate the desired
20 change accordingly. To this end other tactics may be used, for example, a threat to
21 collectively sponsor negative advertising or editorials. In fact, the above technique for
22 presenting a believable threat could be utilized in this case and backed up by an insurance
23 policy which provides in addition to the guaranties that the threatened action will be carried
24 out, for example, by X date unless a particular condition is met, also provides sufficient
25 funds to follow-through accordingly and in one variation further instill the motivation on
26 behalf of the parties by actually insuring the outcome of the litigation (where the threatened
27 (and guaranteed) consequence is sufficiently great that the entity is much rather proved to
28 settle thus reducing exposure significantly).

29 Within the present system SDI provides the framework by which appropriate users and
30 sellers may be matched together. It also enables a methodology by which the user interests

1 are protected through the use of matching of sellers offers to competitive vendors (using
2 iamworthit).

3

4 **2.1.7 Single-user Time-aggregated Purchases**

5

6 Another extension is to propose a special interest-bearing bank account which we set-up for
7 the user which is tied into a wallet or debit/credit card for off-line transactions. We can
8 allow a user agent to bundle its own purchases over time, and use the system of SDI to
9 guarantee future payment to on-line vendors in return to a good current offer. This is novel,
10 because it is like participating in a “discount scheme” without purchasing the right to future
11 discounts up front. A user can still invest money that is earmarked for future purchases with
12 a vendor.

13 We describe a wallet application for this purpose, that allows a user of SDI to commit
14 currency to vendors for future purchases. The time-of-purchase competition system can
15 allow vendors to see the pseudonymous purchasing profiles of users (e.g., over the past 12
16 or 18 months) which are relevant to their respective commercial venues and competitively
17 bid for the opportunity to acquire from that user a “pre-commitment” for similar spending
18 amounts within the same time period going forward (provided that that vendor offers the
19 same/similar categories from which the user had purchased that amount in the past from
20 another vendor(s).

21 The “committed currency” in the user’s account is not actually spent and thus remains
22 interest-bearing for the user, and because the dollars are exclusively and irreversibly
23 earmarked, the actual debit could even occur substantially after the actual purchases
24 ultimately occurs with that vendor. Vendors marketing the service may be able to exclude
25 competitive offers (as is the case with the standard time-of-purchase competition
26 application).

27

28 **2.1.8 Special Vendor Treatment in a Community Dollar Scheme**

29

1 purchases at any of these vendors in the community. SDI can be used to identify optimal
2 combinations of vendors, to best satisfy the customer's needs. Unlike prior art multi-
3 vendor loyalty point systems, in which the selection of vendors is static across all users,
4 this allows the dynamic creation of loyalty schemes. This is possible because of the
5 consumer preferences which can be made available via SDI. In cases where vendors offer
6 some common products it may be possible to redeem value at either vendor's site, or to
7 restrict a purchase to products at only one of vendors. From a data analysis perspective a
8 primary criteria for matching vendors is to identify those which have products predicted
9 to be interesting to the user are metrically close to each other within user (cluster) space
10 (are predicted to be preferred by many of the same customers) AND have a minimal
11 number of products which overlap, thus minimizing collaboration among competitors.

12
13 The actual debiting of a user's account can occur substantially after a commitment to
14 spend money with a vendor, allowing a user to continue to accrue interest. In this multi-
15 vendor mini market the money can be earmarked for that vendor exclusively at the
16 moment of purchase.

17 18 **2.1.9 LEIA Enhanced Time-of-Purchase**

19
20 One application of LEIA involves the integration of the time of purchase competition
21 architecture. In particular, purchasables which a user is browsing with an associated offer
22 price may be either purchasable on-line or off-line. Based upon his/her present physical
23 location, if the user so desires s/he may identify not only other similar items but also that
24 subset of these items sold by bricks and mortar vendors which are physically proximal to the
25 user's current location (thus available for immediate purchase). By introducing location
26 enhanced information to users about their relative similar proximities to off-line
27 purchasables the time of purchase vendor competition model is further enhanced by placing
28 additional competitive pressure upon those vendors selling similar goods exclusively on-line
29 (which are not immediately available to the user). In this environment if an off-line product
30 offer is ultimately accepted by the user, a digital certificate could be securely conveyed to

1 the user or authentication of his/her identity could be established at time the offer was
2 accepted and upon redemption.

3 The Location Enhanced Information Architecture (LEIA) enables vendors in the vicinity
4 of where a user is traveling to deliver a targeted ad or particular personalized offer. Or a
5 user may search or navigate directories on the Web or browse a particular (competitive)
6 item. Regardless of the mode of access, the system filters or identifies purchasables by
7 their direct physical proximity to the user (or the user's intended travel path). If the time
8 of purchase competition system uses location proximity of an item as a parameter for
9 vendors to compete with one another (i. e. providing an additional speed/convenience
10 factor for on-line items with an off-line sales presence) the particular vendor's price
11 which is finally agreed upon and accepted by the user can be securely enforced at the
12 point of sale using a digital coupon which is non-transferable and non-reusable may be
13 down-loaded to the user's device, printed or simply securely transmitted to him/her as a
14 unique code redeemable with user identification at the physical point of sale. These
15 techniques are fully disclosed in co-pending patent application entitled "System for the
16 automatic Determination of Customized Prices and Promotions".

17 **2.1.10 Extended Example: Medical Insurance Application**

18 *Problem statement:*

19 Increasingly, insurance companies are denying claims for medical services. Hospitalized
20 patients frequently have in-patient days denied, in part or totally, and associated services and
21 studies may also be denied. Days of partial or total denial are especially important for non-
22 DRG (Diagnosis Related Group, Non-Medicare) insurance subscribers. Also in-patient and
23 out-patient testing and services are frequently denied at the discretion of insurance
24 administrators. These denials persist despite consistent efforts by hospitals and hospital
25 personnel to properly document the appropriate protocols to alleviate patients' illnesses.
26 Similarly, out-patient medical care in doctors' offices is not infrequently denied, and to the
27 extent of prohibiting effective diagnosis and treatment. At times, hospitals as well as out-
28 patient office personnel do not have the necessary information to know, prior to treatment,
29 which services are eligible for reimbursement, partially because of the multitude of coverage
30 patterns. The providers must frequently appeal non-payments retrospectively, by letters,
31 further documentation and repeated telephone calls, all of which may or may not be

1 effective, apparently depending on the capriciousness and the judgment of the particular
 2 clerk or medical personnel in the carrier's office. These situations have produced enormous
 3 pressures to improve communications by extracting detailed data from providers pertaining
 4 to carriers failing to support appropriate and necessary health care delivery and
 5 documentation of how these insurance issues have caused definitive health problems. This
 6 data will be used to exert pressure on insurance carriers. In view of these major economic
 7 pressures, hospitals and other providers of medical services will definitely be most
 8 cooperative in providing detailed billing and clinical information with appropriate release
 9 consents, or alternatively through deployment and utilization of the proprietary techniques
 10 for privatizing patients' records using a pseudonymous proxy server. This information will
 11 be provided by the billing and medical records departments of the hospital and also, in the
 12 case of hospital employees, from the Human Resources Department of the hospital for
 13 reimbursement for medical services for hospital personnel. The minimal information should
 14 include the billing record for those in-hospital days or the specific services/tests denied and
 15 the discharge summary of that hospitalization. Based on reviews of this material, further
 16 portions of the in-hospital clinical record may be requested. (Additionally, the iamworthit
 17 service provides for the utilization of the above mentioned proxy server for collecting and
 18 maintaining pseudonymous patient data whereby advertising revenues from industry
 19 advertisers accessing/targeting desired user profiles may be shared with the hospital and
 20 potentially also the patient, thus further insisting the hospital to provide this needed patient
 21 data).

22 A compounding problem is the lack of information provided prospectively by insurance
 23 companies to hospitals, physicians and other providers. Subscribers and providers have
 24 the right to know, in detail, the number and kinds of services covered, and the depth of
 25 the insurance coverage including the length of stay. This information is legally
 26 accessible and can be made available not only to the patient, but also to his/her physician.
 27 On exemplary commercial need for the present service by employers who wish to
 28 reinsure their employees with quality coverage.

29 *Solution:*

30 In order to protect the consumer, a solution proposed to address a timely and
 31 overwhelming need on the part of consumers to establish a control data bank,

1 accessible to the consumer, which reveals general and specific instances of
 2 miscarriages by health insurance carriers of their fiduciary obligations to provide
 3 timely payment for essential health care services. Based upon data regarding the
 4 insurance company's stated policy and statistical analysis of the ultimately approved
 5 or disapproved coverage, performance criteria may be determined pertaining to the
 6 percentage likelihood of the insurance company covering certain services and length
 7 of hospital stays for future services to patients. In order to provide a robust and
 8 comprehensive statistical analysis, it is also required that a certain core sample of
 9 non-coverage denied service be collected in order to determine a variety of
 10 correlation criteria which affect the probability of coverage in addition to the identity
 11 of the carrier, also the type of prescribed medical care for associated illness wherein
 12 coverage for hospital stay or treatment is denied. Other correlations are possible, e.g.,
 13 the medical history of the patient, the family history, etc. Of additional importance,
 14 statistics may also be provided which indicate the degree of adherence (integrity) of
 15 an insurer to provide coverage and to provide it with the level of depth of coverage
 16 which was purported by the insurer according to its standard claims policy. The
 17 iamworthit insurance service is accordingly designed with two primary case
 18 objectives:

- 19 1) Provide detailed statistics based on robust historical data regarding realistic
 20 expectations which the patient can anticipate regarding acceptance, denial and quality
 21 on the basis of various types of coverage given the context of the insurance needs and
 22 requirements of the buyer and based upon the specific services and depth of coverage
 23 purported by the insurer and any associated medical information regarding the buyer.
 24 The service could even anticipate, by review of the collective medical statistics the
 25 most likely type of coverage the buyer may need, based on his/her medical record
 26 specifically within the context of the associated determined probability of coverage
 27 and quality thereof for each insurer in view of this medical history of the buyer. In
 28 addition, iamworthit may provide a general informational resource where buyers can
 29 access such statistical information by carrier. Data mining tools may enable the buyer
 30 to extrapolate the desired statistical correlations as to his/her type of profile of
 31 medical or specific features of his/her medical condition or specific type of coverage

1 and thus in light of this robust data better equipping the buyer to determine which
2 insurance provider and associated policy is most appropriate for his/her specific
3 health care coverage needs;

- 4 2) This information may be accessed in real time by the buyer (or for e.g. buyers,
5 employer) while browsing specific insurers' sites, receiving specific offers by
6 insurers (typically requiring certain medical information) or browsing an insurance
7 retail portal. Typically, buyers of the present service will combine their use of the
8 present browsing-based information with the Time-of-Purchase Vendor Competition
9 Services (described below). Additionally, a rating feature may also be provided
10 which enables the buyer to rate (and subsequent buyers to observe the averaged
11 ratings) the various vendors by various types of criteria relevant to the buyer's
12 experience in receiving acceptance, quality and depth of coverage (as well as how
13 well the insurer adhered to promises for coverage as stated in its standard policy for
14 coverage), of course, this feature could be used if for insurance products on-line for
15 later purchase off-line.

16
17 Iamworthit's time-of-purchase vendor competition application to medical insurance (as
18 applies similarly to numerous other e-commerce commercial domains) involves the
19 following protocol:

20 The iamworthit enabled consumer provides specific needs or requirements as to what types
21 of medical insurance products or areas of coverage are desired. These requests may be
22 submitted either to a portal (or "mall") which connects the buyer to an extensive resource of
23 medical insurance providers, or, alternatively, these requests may be submitted to
24 iamworthit which provides its own default portal to these resources (which likely exists and
25 is marketed as an independent marketplace for insurance purchasers on its own behalf).
26 Upon submitting these buyer requests, iamworthit is able to provide numerous competitive
27 insurance providers offering similar insurance products which in this application may be
28 accompanied by certain credentials (such as the buyer's age and many other clinical
29 parameters).

30 Additionally, the pseudonymized digital medical records of the buyer (which may also be
31 requested by an insurer, typically in addition or instead of submitting a questionnaire

1 regarding the buyer's health such as certain pre-existing medical conditions). Typically,
2 patient medical records are a prerequisite to submitting an offer. The insurers may then
3 provide specific offers for their respective products. Additionally, the buyer may browse
4 various insurer's sites, providing pseudonymous medical data to the extent (and type) that it
5 is requested and, specifically, to which providers the buyer desires. As offers from
6 insurance providers are collected (in response to requests and/or browsing interactions), the
7 time of purchase vendor competition is elicited whereby each insurance provider is
8 automatically notified as to the most valuable offer(s) provided by a competitor(s) and are
9 given the opportunity to submit counter offers accordingly. It should be noted that the
10 determination of the value may be in accordance with certain buyer prescribed
11 specifications such as price or certain expert determined criteria which consider either the
12 price in combination with completeness and depth of coverage and, if relevant, the level of
13 satisfaction of the buyer's request criteria.

14 Additionally, iamworthit provides the ability for buyers to share the information of offers
15 they have received in accordance with the associated medical profile and/or response to
16 health related questionnaire information, in order to determine if an offer which is submitted
17 by an insurer inappropriately discriminates against the current buyer in terms of price or
18 depth/comprehensiveness of coverage, or where such discrimination occurs in conjunction
19 with buyer medical data which is prohibited by insurance regulatory guidelines as a valid
20 criteria for such discrimination. Iamworthit's insurer coverage statistics (which the buyer
21 may typically wish to observe in conjunction with the present time of purchase vendor
22 competition functionality) are an additional important value determination factor (which
23 may be factored in as an additional weighting criteria as numerically prescribed by the
24 service (or by the buyer) and/or accordingly for purposes of setting a threshold by which
25 insurers become disqualified as candidates for the buyer's business. In a variation, as is
26 provided by the current model, buyers may instead first receive the historical offer
27 information from other "similar" medically profiled buyers in order to submit a bid to that
28 carrier which offered the most relevant and beneficial offer as well as all of the other
29 competitive insurers. As is also provided according to the current scheme, the present
30 iamworthit time of purchase vendor competition facilities may operate fully autonomously
31 on behalf of the buyer (incorporating, if desired, specific request criteria).

1 Thus, optimal offers are provided to the buyer by a "buyer agent". Additionally, the present
2 service may with some modification be provided to buyers dialing into an 800 number.
3 Typically, in this case, however, (as is also the case in the on-line implementation), if the
4 buyer does not have access to his/her pseudonymous medical records, he/she would
5 typically answer a questionnaire for each provider (which would typically consist of an
6 aggregate questionnaire satisfying the requests of all relevant carriers which could then be
7 submitted automatically, or an agent could be programmed to automatically complete the
8 questionnaire as needed.

9
10 *Commercial Alliances with Regional or National Health Care Organizations –*

11 Because of the tremendous restrictions in providing insurance coverage to organizations
12 which provide health care services, it will be greatly to their advantage for patients and
13 physicians to be able to benefit from more fair and competitive insurance coverage by using
14 time of purchase competition and the associated coverage probability service (providing full
15 disclosure of non-coverage and inequities in delivery of coverage). The primary joint
16 marketing entities which will benefit most significantly from this service and thus will be
17 the strongest commercial allies, include independent physicians, medical clinics, large
18 physician groups, pharmacies and perhaps even pharmaceutical companies, providers of
19 medical supplies, hospitals and home health providers. Of the above entities listed,
20 physician groups, clinics, and hospitals do a tremendous amount of advertising directly to
21 consumers (and this is becoming increasingly true via the internet for physicians in private
22 practice), particularly through the direct advertising medium of the internet (e.g., advertising
23 on their web sites as well as email). These organizations, in light of their direct marketing
24 exposure to the consumer which they provide, would likely become powerful allies in
25 assisting the promotion and exposure of the competitive insurance with full disclosure of
26 claim information service to their consumers.

27 Secondly, health care organizations which market to health care professionals, such as
28 physicians and nurses, including pharmaceutical companies, pharmacies, and home health
29 care providers may indirectly benefit by informing their target customers about the above
30 consumer oriented service.

1 *Quality Ratings of Medical Suppliers*

2 iamworthit's insurance application provides a natural entry into a second commercial

3 application of the service. In a fashion similar to that of Time-of-Purchase Vendor

4 Competition Scheme, for purchasers of insurance products, another useful commercial

5 application of the present service is facilitating the time of purchase vendor competition

6 between medical suppliers and equipment providers. The synergy by which this second

7 commercial application is enhanced is as follows: physicians and hospitals will likely be

8 cooperative allies to the service in providing patient information substantially as required as

9 well as promoting (along with home health care providers) particularly the iamworthit

10 insurer coverage probability service to their consumers. They will also likely utilize such a

11 resource for internal uses and if/when HMO service liability legislation becomes enacted,

12 for legal purposes as well.

13 To the buyer's advantage, it is possible (as in the insurance application, albeit less likely)

14 that unfair or arbitrary price discrimination may be detected and counteracted by

15 iamworthit's data sharing feature. Common buyers in this environment include hospitals,

16 physician clinics, home health care providers and to a lesser extent HMOs and end-buyers,

17 the nature/types of supplies and equipment which tend to be purchased by each of the above

18 categories of buyers are provided in further detail below. As suggested earlier in this

19 description (and in issued patent "Pseudonymous Server for System for Customized

20 Electronic Identification for Desirable Objects"), a vertical portal as the one described, may

21 be personalized based upon the user profile as disclosed upon accessing that portal in

22 general, at the item level as within a category (or in conjunction with search results).

23 Additionally, vendors may be equipped with price discrimination tools and are in fact highly

24 motivated to form a coalition by entrusting their data and counter offer responses to

25 incoming time of purchase vendor competition offers and bids to SDI (or at least an industry

26 specific counterpart of SDI) for medical suppliers in order to provide limits or "price

27 ceilings" for any given buyer and associated profile and condition relating to competitive

28 offers or bids.

29 As in the case of almost any category of purchaser, particularly the larger entities, such as,

30 unions, hospital systems and large physician organizations, additional bargaining leverage

31 can be applied to suppliers because of the volume of supplies which can be purchased. The

1 iamworthit model is particularly also likely to facilitate the buyer's leverage when
2 negotiating long-term contracts with suppliers through the buyer-buyer data sharing and real
3 time competition/bidding facilities which iamworthit provides.

4 Finally, it is worthy to note that as in the insurance application, a rating feature may also be
5 provided which enables the buyer to rate (and subsequent buyers to observe the averaged
6 ratings) the various vendors by various types of criteria such as quality of products,
7 promptness of delivery, quality of service (or as in medical equipment, quality and
8 promptness of technical assistance/repairs), availability of desired inventory etc.

9 There are a variety of types of health care organizations which could directly benefit from
10 the time of purchase competition and quality rating service of medical suppliers which are
11 described below:

12 (a). Physicians Clinics and Large Physicians Groups - Typical types of medical supplies
13 and equipment which these organizations will purchase using the iamworthit medical
14 supplies application include: equipment such as X-ray equipment, ultrasound units, EKG
15 units, EEG units, computer and other office hardware, chemical analyzer units, clinical
16 supplies, office furnishings, pharmaceuticals, etc.

17 (b). Hospitals and Hospital Systems - Typical medical supplies and equipment which
18 hospitals will accordingly purchase include: MRI units, Cat Scan Units, X-ray units
19 (including units for special studies and mammograms), radiation therapy units, stress-testing
20 equipment, hematological and chemical testing equipment, surgical supplies, prosthetic
21 supplies, patient-care supplies for medical and surgical floors, IV solutions and
22 pharmaceuticals, physical plant operational and replacement equipment and supplies, etc.

23 c). HMOs - Currently, a certain percentage of patient health plans include coverage for
24 pharmaceutical products. These products are typically provided through contract with
25 centralized pharmaceutical suppliers. It is accordingly reasonable to suspect that many of
26 these HMOs will invite the opportunity to provide these supplies through a more open
27 dynamic marketplace environment as is provided by the iamworthit service.

28 (d). Union Directed Health Care Organizations - Union clinics typically purchase
29 equipment and supplies, similar to private clinics or large physicians offices, as described
30 above. Union private health care organizations would be compelled to incorporate the

1 iamworthit medical supplies application as it similarly could apply to all of the above
2 entities.

3 (e). Home Health Care Market - Equipment and supplies which are, in part, similar to large
4 physicians clinics as well as some hospital - type supplies, used to provide out-patient care
5 and maintenance services.

6 *Patient Medical Records*

7 The secure data interchange provides a secure and privacy protected storage and retrieval
8 architecture in which buyers (in this case patients) are able to control the use of their
9 personal profiles (in this case medical record information). Herein, we propose an
10 alternative model which is based upon the assumption and belief that purveyors of this
11 medical information, including hospitals, physician's offices/clinics, labor unions, (to some
12 extent) HMOs are legally permitted to disclose and use their information for advertising
13 purposes so long as a method is provided which insures that their information will not be
14 associated with any identifying information about the patient (name, address, social security
15 number, etc.). If regulatory constraints and/or the purveyor of this information prohibits the
16 use of this information for the above purpose, patient permission could be acquired by
17 patient signature on all appropriate consent forms at the time of admission and directly on
18 the admission form.

19 In one variation, software may be provided to the organization which enables the
20 organization to operate their own pseudonymous proxy server, thus enabling access by
21 advertisers to the pseudonymized patient records. Email and/or telephony-based
22 pseudonymous communications could be readily provided. For direct mail, the address
23 information maintained by the hospital would mandate the hospital to provide direct mail
24 advertising to the patients or to entrust this operation to a pseudonymous physical mail
25 service such as SDI. In the (much) preferred variation, a trusted pseudonymous proxy
26 server operator maintains the patient information on an external database (this would
27 perhaps be SDI or an SDI-like service) in which all of the above pseudonymous
28 communication media could be provided along with pseudonymous direct mail services.
29 For the SDI data entrusted variation, the preferred business model to provide incentives
30 to these organizations to cooperatively provide this data would involve SDI's privacy
31 policy enforcement capabilities in combination with a revenue sharing model in which

1 revenues from advertisers such as pharmaceutical companies, pharmacies, clinics,
2 HMOs, hospitals (where the purveyor of the data is not a provider of the same
3 competitive services). In the case in which HMOs are the advertisers, the HMO would
4 identify patients who are currently subscribed to competing HMOs and (typically) have a
5 medical history indicative of a good risk to the HMO. The iamworthit insurance
6 competition scheme, as above described, could be implemented within the context of this
7 current solicitation scheme in order to elicit time of purchase competition among HMOs
8 upon the submission of an offer by the original HMO to the pseudonymous patient.

9
10 advertising purposes so long as a method is provided which insures that their information
11 will not be associated with any identifying information about the patient (name, address,
12 social security number, etc.). If regulatory constraints and/or the purveyor of this
13 information prohibits the use of this information for the above purpose, patient permission
14 could be acquired by patient signature on all appropriate consent forms at the time of
15 admission and directly on the admission form.

16 In one variation, software may be provided to the organization which enables the
17 organization to operate their own pseudonymous proxy server, thus enabling access by
18 advertisers to the pseudonymized patient records. Email and/or telephony-based
19 pseudonymous communications could be readily provided. For direct mail, the address
20 information maintained by the hospital would mandate the hospital to provide direct mail
21 advertising to the patients or to entrust this operation to a pseudonymous physical mail
22 service such as SDI. In the (much) preferred variation, a trusted pseudonymous proxy
23 server operator maintains the patient information on an external database (this would
24 perhaps be SDI or an SDI-like service) in which all of the above pseudonymous
25 communication media could be provided along with pseudonymous direct mail services.
26 For the SDI data entrusted variation, the preferred business model to provide incentives to
27 these organizations to cooperatively provide this data would involve SDI's privacy policy
28 enforcement capabilities in combination with a revenue sharing model in which revenues
29 from advertisers such as pharmaceutical companies, pharmacies, clinics, HMOs, hospitals
30 (where the purveyor of the data is not a provider of the same competitive services). In the
31 case in which HMOs are the advertisers, the HMO would identify patients who are currently

1 subscribed to competing HMOs and (typically) have a medical history indicative of a good
2 risk to the HMO.

3 The iamworthit insurance competition scheme, as above described, could be implemented
4 within the context of this current solicitation scheme in order to elicit time of purchase
5 competition among HMOs upon the submission of an offer by the original HMO to the
6 pseudonymous patient.

7 2.2 Matchmaking

9 The user-centric SDI model allows users to provide personal information on a carefully
10 controlled basis to vendors and other users. Furthermore, vendors can implement rules that
11 personalize the information, products, and service provided to users—on the basis of
12 personal information that they receive from users directly, or have acquired about users. The
13 key property of “matchmaking” applications is that information exchange is bi-directional,
14 i.e. a match is not computed on the basis of static information within the system, but may
15 involve additional information exchange between agents, and/or special actions by one or
16 more agents.

17 Client-level SDI proxies can act as autonomous agents in an architectural variation of SDI,
18 where the “client-level proxy” is co-located with a (physically) mobile user, for example on
19 a palm-held computer or head-up display. In a “match-making” application the goal of these
20 SDI proxy agents is to find other agents with desired profiles/synergies, with the view to
21 taking an action or exchanging a particular type of information. We describe many
22 applications, for example: to matching buyers with a virtual sales force, to a market for
23 expert opinion, to groupware systems for virtual team working, and to an investment
24 system.

25 We can allow agents to communicate *anonymously, without revealing (even*
26 *pseudonymous) identities*. In addition, we can prevent one agent receiving profile
27 information about another agent unless authorization is provided within the rules of data
28 release associated with the agent’s profile. Implicit authorization occurs when a requestor
29 can present certificates to verify that it has required attributes to access particular
30 information. Explicit authorization occurs when a requestee provides direct authorization
31 to a particular part of a user’s pseudonym.

1
2 Essentially there is bi-directional information filtering: the requestor agent will only present
3 certain information to the user, information that is relevant; and the requestee will only
4 provide information when a request is judged to be legitimate. Information exchange
5 between agents occurs as part of a multi-step negotiation, until both parties can agree on
6 terms for either a physical meeting (or execution of a deal), or further pseudonymous
7 exchange of information or cooperation.

8 In match-making applications the role of the central SDI query execution engine is to search
9 continuously for "matches" between agents, based on agent profiles. When a match occurs
10 additional information exchange may occur automatically between agents, until finally an
11 introduction is made or an action executed by one (or both) of the agents. Decisions about
12 what information to exchange are made on the basis of both static and dynamic profile
13 attributes, e.g. standard (historic) profile information, current behavior, current location, and
14 recent activity. Also mode of behavior e.g., social, business, leisure as detected by the
15 user's calendar, time, content in recent conversations, type of users recently interacted with,
16 etc. LEIA style-behavior attributes can be used to automatically decide on the relevance of
17 new virtual tag information. A requestee might also demand certain credentials to indicate
18 the lack of negative reputation marks, for example that an interaction with the user has never
19 received a bad rating (see Section 2.2.3). Perhaps a third-party could be used to determine
20 whether the user's know each other (e.g. www.sixdegrees.com). When a requestee denies a
21 request for information, it may instead provide criteria for data releases. A requestor can
22 respond with a different information request, or a subset of required credentials. Finally, the
23 agents might agree on terms of negotiation and conditions can be anonymously fixed.

24 There are (at least 5 levels) of information disclosure: indicate to another user interest;
25 release profile information; disclose communication; start a correspondence session;
26 schedule a meeting/strike a deal, take some other action, etc. The end-result of information
27 exchange could be an agreement to calendar a meeting for some future time and place; and
28 absolute, or pseudonymous revelation of identity.

29 An initial implementation of the data-release policies might allow only manual definitions.
30 However, after an initial "beta testing" phase, a data mining suite could be used to cluster
31 users and generate exemplar data release and data request policies. A system can provide

default settings for users, and recommend setting based on users with similar profiles. The user can further fine-tune the rules. Automatic feedback techniques can be useful to adjust rules, for example—when a user is especially receptive to particular type of introduction then make such introductions more likely in the future. An intelligent interface system might also suggest refinements to the rules, to automatically cover “patches” where the user currently controls interactions. For example, some rules may be based upon a certain confidence threshold, which the system suggests or refines. The user may accept or adjust the threshold as desired. S/he may wish to provide settings as to which types of autonomous actions the user wishes to authorize updates to (based upon his/her own behavior or as a default users who are very similar) and the user wishes to authorize the recommended action, or allow the system to fully automate the activation of the recommended actions (or do so only if the confidence threshold is above a certain level). Or for example bypass user authorization of a recommended action if it falls within a certain margin of confidence below the normal confidence threshold for activating full automation if the user is performing another important task , e.g., in a meeting, engaged in an important phone conversation

2.2.1 Resolution Credentials

Resolution credentials from third parties can play an important role in agent-agent interactions. We can prove the absence of a quality attribute or behavior (which is often of a negative nature) relating to an individual and is submitted by a third party and typically must be issued on a periodic basis in order to maintain currency. A few simple examples of resolution credentials which may be of interest to users (credentials which users may commonly request as a precondition to requesting or accepting requests to be introduced or initiate communication with an outside unknown third party) include:

- 1) For business associations: are in good business standing, e.g., have not attempted to defraud other users in the course of common business practices. Or maintain sufficient funds in one’s account to perform business activities (as represented by the user).
- 2) For business interactions or social interactions: are in good standing with the law.
- 3) For social interactions: have not been accused by other individuals of inappropriate or antisocial behavior.

1
2 Some standard credentials which may be of interest to many users, and which may (as with
3 resolution credentials) be incorporated with the standard settings of the user's data request
4 policy as herein described. A few examples are cited (among countless potential others):
5 profession, awards, honors, alma mater, e. g., Harvard graduate, doctorate degree, etc. In
6 accordance with the parent issued patent US Patent #5754938, various credential issuers are
7 provided for issuing standard and resolution credentials to individuals. Thus certain entities
8 may be entrusted with "legitimate authority" to validate and submit credentials which are
9 issued to the appropriate individuals. If a resolution credential is not issued (or not renewed)
10 an adjudicating third party is provided which has access rights to both of the parties is
11 provided to resolve resulting disputes (from the subject user). The present invention
12 describes how credentials can be issued to users pseudonymously.

13 There are a variety of rules which a user's data disclosure policy and data request policy
14 may contain, to control what if any attributes are released, and what credentials are
15 required. A data request policy may state a rule for explicitly notifying the user if a
16 particular resolution credential (e.g., indicative of a serious problem or concern) cannot
17 be presented in response to the user's disclosure request. We allow initial information
18 exchange to be anonymous, such that information that is released as preconditions for
19 release of further information is not useful. Similarly, so long as initial encounters are
20 anonymous there is no need to withhold information about them from the user.

21

22 Some users may not wish to disclose specific information about themselves via these
23 standard credentials but instead certain "extracted" more general information may be
24 provided about themselves. For example, instead of a "Harvard grad or Ph.D." there may be,
25 for example, credentials indicating "intellectual" or "prominent intellectual". Or instead of
26 indicating an individual's wealth or value of assets, the credential may indicate "wealthy" or
27 "very wealthy" (typically, depending upon user's wishes this latter credential should also be
28 withheld during initial introductions or subject to some fairly stringent conditional criteria
29 from the other party) and instead replaced with an even more general credential e.g.,
30 "prominent" or "influential citizen"). Similarly, an individual's exact profession or scope of
31 work may not be fully disclosed initially but rather a more general definition of his/her

1 behavior, and present temporal interest of a user and/or third party for purposes of
2 employing the user's data disclosure and data request policies.

3 Credentials can allow users to identify other users that may pose a threat. This identification
4 may be provided vis-a-vie resolution credentials and/or rating (by third parties). e.g. a user
5 has not engaged in any serious criminal activity, physically harmed another person, or
6 interacted with other individuals who are unable to produce these resolution credentials.

7 Other credentials may specify the nature of an infringement, and its context and severity
8 (e.g. what was the context of a physical assault? Was it performed during a bar brawl,
9 against a friend, a boss, an elderly person, a child, a family member – or at work? In this
10 case, the user agent may, for example, bring to the attention of a prospective employer that
11 the user could not present a credential indicating that they had not previously harmed or
12 threatened a former employer. Was it minor or severe? Also, if such individuals (lacking,
13 for example, resolution credential proving the absence of having committed armed robbery)
14 are (or come) within a certain proximity of a user, the user may wish to program his/her user
15 agent to notify the user. The same would, of course, apply to a store clerk regarding
16 customers of this sort or to baggage security personnel at an airport. Or, highway patrollers
17 may be interested (e.g., on certain stretches of highway) in being made aware of vehicles
18 and their locations whose agents are unable to provide a resolution credential proving the
19 absence of a drug conviction.

20 In another application (in accordance with the auto insurance risk determination methods
21 described in co-pending patent application entitled "Applications for Location Enhanced
22 Information Architecture" [INSERT PATENT OFFICE NO]), an on-board computing
23 device within a user's automobile could identify another automobile lacking, for example, a
24 resolution credential for safe driving. i.e. the on-board user agent continuously polls agents
25 in other cars for a "safe driving" credential, and if it fails to receive such a credential it
26 issues a warning to the user. As an extension, this location data could be converted into a
27 dynamic 2-D rendering upon the user's windshield (using heads up display technology) in
28 order to thus superimpose a persistent flagging or highlighting of that particular automobile
29 from the driver's visual perspective. Pedestrians or law enforcement officials (for example)
30 could also receive instant notification. As is described in LEIA, a roaming cellular
31 connection, or GPS, is not essential for providing a user identifier. For example, optically-

1 based biometric identification techniques such as iris scanning or combined iris/facial
2 identification techniques may be used among other potential inputs as well. Users will be
3 reluctant to release location/time data, even anonymously, when suspicious behavior can be
4 inferred—probably subjectively.

5
6 In the case of suspicious behavior we can provide access to law enforcement officials,
7 through seizure of the decryption key for that data (which includes his/her physical location
8 information) and any additional profile data which is considered of immediate critical
9 relevance to the suspect (or prospective) infraction. Such cryptographic techniques for key
10 seizure from a key escrow are well covered in the literature [FRED, WHAT ARE THE
11 REFERENCES?]. There may also be certain circumstances in which key seizure may be
12 required after the fact (at some time in the future). For example, if/when certain even
13 moderately “suspicious” behavior patterns are detected, it may be possible for the SDI data
14 warehouse to preserve a comprehensive record of that information (and perhaps the record
15 of that user which precedes and follows that period of interest). Thus preserving evidence
16 which may later prove useful in contributing evidence towards a conviction, acquittal, e.g.,
17 proving that a user was not at a particular location/time. A record containing more detailed
18 segments of a user with a proven negative or questionable history may be preserved and
19 general location/time features may be abstracted for the remaining portions of the record
20 (thus compressing the record substantially). This may be performed for regular individuals
21 as well, thus retaining key relevant features while discarding the majority of the record
22 which is irrelevant or redundant.

23 Within a location enhanced context, unless the prescribed range of “proximity” to the user is
24 quite large, securely protecting the user’s identity from malicious third party collusion (for
25 purposes of combining unique pseudonyms and/or exchanging data that has been released
26 and entrusted to them) is a harder problem. The system could (most obviously) assume data
27 exchange between the parties will occur and limit the combined disclosure to only that of
28 the most data restricted user in a given location/time domain. The system could
29 alternatively, perhaps “space apart” the number of users within a given location/time context
30 who can access more “restricted” user data (of course the problem goes away if all the
31 discloseses have similar disclosure restrictions by that user). The use of resolution credentials

1 within the context of a location enhanced (physical) environment will be hard to implement
2 practically until there are powerful and miniature mobile devices. Automobiles may be an
3 exception, as may be technology which enforces the disclosure of the physical presence of a
4 user agent (resident in a device) to other user agents within the networked environment.
5 Continuation of the Addition to Autonomous Information Disclosure to Third Parties
6
7 The above description suggests the use of a statistical model which can take implicit
8 passive data about the complete set of variables surrounding a particular instance in order
9 to determine whether, and if so, what particular data the user is most likely to wish to
10 disclose, such as the identity of the individual, the profile of the individual (e.g., what
11 other users s/he is most similar to that have also interacted with the user), the time,
12 location, inferred activities and content based attention focus (which may utilize the
13 above variables in addition to passive monitoring of spoken content and facts surrounding
14 conversations). A few concrete examples may be worthwhile suggesting .The system ,
15 for example, determines that a physical introduction between two individuals may be of
16 potential interest to them and compatible with their schedules. In fact, they are able to
17 briefly connect (perhaps pseudonymously initially). Mr. A suggests to Mr. B that a lunch
18 meeting may be a potential interest (assume that at least one of their calendaring agents
19 does not have them committed to a fixed schedule prior to lunch time and one of these
20 individuals is Mr. B). Mr. B replies also indicating a potential interest, however, that he
21 has recently dropped in at a friend's house and wishes to stay at least 20 minutes longer.
22 Based upon the schedule of Mr. A, SDI receives data from the navigational agents of Mr.
23 A and Mr. B which take into account the distances, drive time (based upon the preferred
24 routes, the average speeds which Mr. A and Mr. B tend to travel for similar types of
25 interests, and the café or restaurant which is closest to the point at which they rather are
26 likely to cross. Mr. B , may state the constraint to his agent, i.e., the 20 minute minimum
27 period prior to departure. Or the agent may interpret the accurate interpretation of this
28 fact from his/her conversation with Mr. A. Or if the meaning or the words are unclear, it
29 could even automatically reiterate its interpretation. The system may tell user Mr. B that
30 in order to meet up with Mr. A (based on Mr. A's schedule) he must leave his friend's
31 house in 10 minutes and forego the regular stop at the convenience store. Mr. C may be

1 another potentially interesting individual for lunch, however, Mr. B's agent doesn't know
2 all of the facts about how Mr. B may be willing to conform his schedule to meet Mr. C
3 with regards to changing or rearranging a certain portion of his schedule, and/or replacing
4 this portion of the schedule with another activity or engagement (an original activity
5 which may now be in conflict due to the drive home and physical direction to visit Mr. C.
6 Mr. B's agent thus identifies and conveys to Mr. B the salient changes to his schedule and
7 the travel route within this possible alternative plan to meet Mr. C instead. Thus before
8 Mr. A commits to Mr. B about lunch he has details about all of his feasible options for a
9 lunch engagement. Mr. B may even state to his agent that that stop at the convenience
10 store is a relatively fixed constraint (that it is important) and given that constraint what
11 are the available options for a lunch meeting as such.

12
13 Note (as suggested) that at any time the system identifies within a certain probability
14 threshold for that user that there may be a certain engagement (or other activity) which
15 could be scheduled (or replace another schedule activity) the system:

16
17 1). Determines the predicted interest-level of that potential engagement or activity (or if it
18 may involve replacing more than one activity/engagement, determine that the predicted
19 interest of the combination thereof and/or it may enable other activities/engagements to
20 be scheduled which may be replacements for others (estimate the predicted interest of the
21 activities/engagements collectively comparing one scenario to the other). There may be
22 other factors affecting user interest in one scenario versus the other, e.g. the distance or
23 travel time with which it is associated. The system may also be uncertain regarding
24 certain important determining variables, e. g. the profile of the user or even the present
25 mood or activity or contents focus is not adequately determined, or the nature of the
26 proposed activity or engagement is not sufficiently constructed (or robust) within the
27 user's profile. In such cases, the user's agent can also ask the user direct questions in this
28 regard using a natural language speech interface and (ideally) at the opportune moments
29 in which the determination of certain facts as these within the user's present user profile
30 would provide the ability to identify or dismiss for the user certain opportunities as they
31 occur.

1

2 In light of the present context of the situation and potential opportunities, the questions
3 may be selectively presented in order of priority and are typically related to the attributes
4 characterizing present opportunities/individuals which are either of predicted interest or
5 of unknown predicted interest/relevance but may possibly also include potentially
6 predicted relevant questions about the present frame of mind of the user, such as what the
7 user's present activities are (i .e., including work , leisure, etc.) mood, present focus on
8 what kind(s) of content, etc.

9 In the above example, of course, Mr. A and Mr. B could instead be established friends or
10 colleagues in which case keeping secret their present physical locations may be
11 considerably less important but not necessarily so. Revealing informational details about
12 location, action, attention focus to another third party while completely leveraging this
13 information to provide optimal scheduling efficiency is all possible with SDI and many
14 individuals may wish to not disclose it in this way (as a default) even if the disclosees are
15 relatively trustworthy.

16

17

18 Complete robust functionality coupled with complete user privacy as provided above
19 through the user's agent provides a compelling motivation for users to maintain much of
20 the information, regarding their present and general location and activity related data,
21 private in the majority of cases of user interaction and associated schedule coordination.

22 If Mr. B were to pose different relevant queries to Mr. A's agent while determining
23 whether he could be interested in meeting, for example, where/when their paths may
24 cross based upon different times and/or points of departure, the specific location and
25 travel information regarding Mr. A. may still be fuzzed as the meeting time at the
26 associated place of rendezvous does not disclose to the other party when the other will
27 arrive, only that time when both will be there at the same time(for those rare cases where
28 even this information may disclose likely specific location and intended time of
29 departure, we can apply randomization techniques in order to assure privacy with
30 relatively minor negative impact upon scheduling efficiency. Nevertheless, how much
31 privacy versus efficiency is desired is entirely within the control of the user.

1 individual and may have professional relevance on the other individual (the latter two
2 examples which may be relevant to say the professional profile of that individual , etc.).
3 In addition, the individual(s) may be made aware of the event disclosure policies set forth
4 by the user relating to them specifically (if desired) and this individual(s) can further
5 provide filtering rules which allow all or a portion of that information to actually reach
6 the user. In another example, the user may have met or known the individual at some
7 time previously and the venue and context of the meeting circumstances and relevant user
8 profile information to that context are known and disclosed by both parties. The user
9 then wishes to re-establish contact perhaps defining the reason for the request and the
10 individual can have predefined rules or acceptance or rejection of the request directly.
11 In a final example, virtual tags are provided containing comments and annotations
12 regarding a certain physical object (which could be anything from a tourist attraction to
13 the physical site of a recent hot news event to a useful annotation critiquing a bricks and
14 mortar vendor). An individual whose profile suggests s/he is knowledgeable (or had
15 provided the relevant annotation of interest) about that object would allow this relevant
16 information to be conveyed to other individuals who have perhaps an interest in that
17 information (generally or at that moment).
18 The relevant individual (if s/he is also willing) may be contacted and solicited to meet if
19 their calendaring agents can appropriately coordinate the meeting. It may be either
20 context specific, e.g., both individuals are physically local (or plan to be local to each
21 other) and the physical object or site of interest or simply (generally) local to each other.
22
23 It is perhaps worthy to note that the present methodology could certainly be extended to
24 include the application to rules dictating reachability conditions of users bu other users via
25 standard telephone. The application to cellular telephony is of particular interest in its
26 ability to communicate to the user in a variety of modalities (voice, notification, instant
27 messaging, Web content). In addition, similar rules may be adapted in this scenario with
28 perhaps minor modifications. For example, under what conditions may the user be reached
29 by another user. E.g., what is the nature of the relationship between those individuals? If it
30 is not known, inferences from the context of previous conversations may be determined.
31 Does the user typically accept calls (or other communications) from the other user under

1 similar circumstances and/or times. If this data is not available are they accepted by other
2 users similar to that user according to a key attribute(s) or does the other user accept
3 communications from the user under similar circumstances as they present communication
4 attempt (as this may suggest similar reciprocal acceptable reachability conditions as that of
5 the other user. Again, LEIA suggesting location movements and associated measurable
6 behaviors as sequential patterns and as formation of time may be important. This detailed
7 statistical model of the user could be uiseful in learning the context of performing certain
8 autonomous actions for the benefit and convenience of the user and in the reachability of the
9 user by other users, e.g., in what communications mode is the user interested in
10 communicating, e.g., speech to text-based instant messaging or standard telephony, does the
11 user mind being interrupted with phone ringers, (or vibration mode) or ringer-off mode).
12 Previous similar communications contexts may be useful in determining inferred rules
13 which may be presently appropriate (e.g., was the spoken content of the previous
14 communication indicative of acceptance or annoyance of the communication with the other
15 user within a similar associated context . Eg., this may suggest whether and in what
16 contextual conditions of both accessor and accessee the accessor's call may be patched
17 through (or re-forwarded if initially accessing another number or telephone of another likely
18 venue where the user may be believed to be present.

19

20 Exemplifications of The Underlying Statistical and Rule-Based Intelligence Used in 21 Autonomous and Semi-Autonomous Release of Personal Data Release, Matching and 22 Reachability by Other Users

23

24 The above section "Location Enhanced SDI System relating to Smart Home and Office
25 techniques provides a relevant platform for integrating intelligence into a multi-device
26 environment. Mobile user intelligence is a sub-component of this broader multi-device
27 (or "ubiquitous computing") application-level intelligence platform. In view of the
28 present application framework, it is a very intriguing and challenging problem to provide
29 a statistical framework which is able to allow the user agent to make appropriate
30 inferences regarding users who generally, though imprecisely, fit a certain profile which
31 may suggest the activation of a rule (i.e., a "fuzzy rule") such as requesting further

1 information disclosure from that user, identifying that individual to the present user or
2 more sensitive types of actions such as if the requested information from the other user
3 requires some sort of a fee or the other user, in turn, requests certain information
4 disclosure from the current user (which may be considered at least moderately sensitive).
5 It is clearly extremely advantageous to entirely automate as many of these types of
6 information exchange and notification related actions as possible in order to enable the
7 system to function seamlessly and intuitively to the user's present wants and needs. In a
8 general sense, it is reasonable to suggest that the key to minimizing active user approval
9 of rule based recommendations is the ability to accurately infer present user actions and
10 behavior, thus allowing the system to more accurately infer the present user mind set,
11 activity related focus, perhaps even the mood which the user is presently experiencing
12 unless the nature of the disclosure is extremely sensitive or otherwise extremely
13 important to the user. Many of these types of actions could be performed without active
14 user approval and authorization. There are a couple of approaches which can be applied
15 which may be able to limit the uncertainty of the system, possibly alleviating before
16 hand, inappropriate or unwanted actions or failure to perform certain desired actions. For
17 example, it might be possible to perform data analysis on very large data sets of users
18 who share very similar profiles and to perform very similar actions and behaviors in all
19 aspects of their physical interactions with the real world. The recommended actions
20 provided by the system may also be used to provide the user with a natural language
21 interface which identifies to the user those types of circumstances and conditions in
22 which there is an inherent lack of adequate statistics from previous user experience to
23 prove beyond a reasonable level of confidence the action desired by the user. Or the
24 system is unable to identify with a high statistical confidence level a most appropriate
25 action based on the collective statistics of other similar users and conditions (the system
26 may helpfully suggest those which are possibly the most prevalent. However, in other
27 cases the actions (or combinations thereof) of the user are unique enough that the system
28 lacks sufficient statistical data, the likely appropriate actions are sharply divided between
29 or among different possible actions or (similarly) there are no strong statistical
30 correlations which can be established (perhaps because the key factors with which the
31 desired actions are most influenced cannot be captured due to the limited input

1 constraints to the present system, e.g., precise accounting of present activities or focus
2 (even such things as the likely emotions or present attention focus of the individual for
3 which there may be a lack of sufficient measurable clues visa-vie the suggested input
4 modalities. The natural language interface (perhaps an application for a language
5 generation system) could communicate to the user its interpretation of his/her behavior
6 and the statistical correlations with recommended autonomous actions. E.g., if the user
7 is in his/her home study between 9-11PM, she/he is likely to be engaged in work relative
8 to a research publication thus an invitation to other colleagues in this area will be
9 extended and/or communication requests accepted .to contact the user accordingly under
10 these conditions. The system may detect similar conditions in which the user may reject
11 requests to correspond by such individuals under these conditions, e.g., if the identity or
12 identified context (or even context inferred from the identity) is inappropriate for that
13 particular context of user activity. The user may provide another annotation to the rule
14 which says that if she/he has recently corresponded with another individual(s) at that
15 time/location, chances are that she/he is, in fact, in a social related mind set. As
16 indicated, there are numerous potential inputs to the system, which could be considered
17 in predicting what agent-mediated actions are, in fact, appropriate. The techniques of the
18 parent patent application, if applied in this context, would allow us to make strong
19 inferences about the particular mind set, reflecting present interests or preferences which
20 the user is likely to be receptive to, presently based upon such clues as who the user is
21 presently interacting with, the content profiles of the present real-time dialogues, e. g,
22 typed or spoken through a communications network, (or simply passively collected off-
23 line), the object profile of documents or web pages being interacted with. Activities may
24 further be inferred by such indicators such as the particular physical location of the user,
25 visa-vie LEIA, the particular sequence of location specific movement patterns, the
26 particular types of devices with which the user is interacting)(and perhaps the nature
27 and/or sequence of those interactions). More complex rules are certainly also
28 conceivable, such, as the user will not interact, transact except under certain specific
29 conditions, such as a commercial opportunity if the user fits a certain stereotype, i.e.,
30 similar to those previously interacted with for the nature or content of the proposed
31 discussion, in light of the type of user, may be acceptable under certain circumstances,

1 locations or mind sets experienced by the user. E.g., a young female in her twenties may
2 not wish to be approached by 40 year old males unless the context of the interaction were
3 purely business, e.g., a professional, or commercial or sales opportunity thus if she were
4 in a purely social mindset such requests to interact would be denied. Likewise a busy
5 executive in his/her 30's may deny such requests from the young female unless her
6 request was purely of a social nature and he was not overly time constrained. There may
7 be certain "professionally" prospective correspondents, however, which he may prefer
8 over others based upon some additional interesting criteria as well such as other
9 business/professional areas of commonality, common interest areas, even socially
10 compatible or interest criteria.

11 Of course, SDI is able to extrapolate extremely sensitive features regarding interest
12 preference and context which may establish a basis where two individuals or more
13 individuals, to be mutually identified to one another and/or communicate accordingly or
14 to not be mutually identified and/or accordingly communicate (SDI may even establish
15 the basis for these mutually beneficial interactions even if revelation of this basis by one
16 or more parties by the other is not permitted within the privacy policy of one or both of
17 the parties). Of course, SDI may frequently identify individuals whose personal data
18 revelation requires certain minimal requested personal data from the other party as a
19 condition to such revelation (which SDI is capable of fully automating on behalf of both
20 parties) or the user may request certain information and make a personal judgment in real
21 time as to whether further information about him/herself should be revealed and/or
22 whether introduction or communication is desirable. In these cases it may be difficult to
23 always successfully achieve an introduction when appropriate as the user is unable to
24 guarantee

- 25
- 26 a). The revelation of certain information to the other use if that user reveals certain
27 information accordingly, or
- 28
- 29 b). Is willing to agree to be introduced if certain desired information is revealed by the
30 other party (often busy executives have absolutely no time to disrupt their busy schedules
31 to discuss certain opportunities when there is enough revealed about those opportunities

to justify the correspondence. However, the nature of the opportunity is too sensitive a piece of information for the disclosing party to be able to reveal without certain guarantees that there is sufficient interest for the recipient to discuss the potential opportunity further. This quandry unquestionably poses a huge challenge, however, it can be somewhat overcome by:

- a). SDI's ability to keep track of most, if not all activities, behaviors and correspondences, particularly acceptance or rejection of those SDI pre-qualified recommended interactions,
- b). Detailed information provided by the individual or entity to SDI as to the nature of the various types of individuals, entities and potential opportunities to conduct business which the individual is able to provide to SDI. SDI itself may apply human experts for the task of interpreting the compatibility of certain requests with the access correspondence policy of the individual, provide a rules based system (which can also learn with fuzzy rules) with which to automate some of the decision or, at least automate the "definitive" exclusions).

2.2.2 Applications

User and seller information - as detailed in the parent issued patent, specific details of what users and sellers may be looking to buy or sell respectively may be used to suggest the basis for a potential commercial transaction. The transaction may be large (but not necessarily so, e. g., real-estate, private investment in a small business or public stock). If a physical or on-line interaction with the other party is warranted (e. g., for larger commercial transactions), as is suggested later as in the present description, users may identify other users which form the most relevant "match" with their interest. At this point the agents can check for credentials, and then either communicate or calendar a meeting. Similarly, the agents may find the "best" match of users who happened to be physically proximal to the user at that particular time, or at some future time(s)/location(s)

1 which is mutually compatible (similar applications are suggested for matching
2 sales persons with prospective clients, identifying experts to work (individually or
3 collaboratively) on a particular project or problem, to answer a question of an
4 appropriate specialized nature to their area of expert knowledge.) The parent
5 issued patent suggests at a general level these commercial applications. An
6 additional feature described therein involves the use of a decision tree called
7 “Rapid profiling” which can be used in the present context to identify from the
8 most common needs of users and “goods” of sellers in general and the known
9 profile data about each user and seller individually, a list of questions for each
10 party which most briefly and efficiently determines the complete user/seller
11 profile of each party individually.

12
13 Social Interests Profile Information—The parent issued patent also suggests the
14 present application at a general level. For a dating application, users may be
15 matched on the basis of their common interests/preferences and perhaps on the
16 basis of certain information reflecting personality, social or cultural
17 behavior/affinities or psychological attributes. On the other hand, for purposes of
18 meeting casual acquaintances, users may be interested in another user who shares
19 the above characteristics as well as someone who has recently shared similar
20 experiences and/or personal challenges.

21
22 Professional Information/Qualifications - As in the application of matching users and
23 sellers, a description of a user’s needs or situation with relation to various
24 professional services may be provided as additional data about the user.
25 Examples may include: (as above) medical data, professional or business history
26 (as well as legal history) which may be of interest to law firms, accounting firms
27 or various business consultants. Personal, family or emotional difficulties may be
28 of interest to psychologists or family counselors. Again, users may submit this
29 information as a query for prospective matches, or they may be pseudonymous
30 queries or automatically matched in accordance with criteria specified by the
31 professional. The issued parent patent application also lists additional

applications, which could as well be relevant within the usage context of virtual tags.

Employer/Employee Information - An employer may post a description as part of his/her virtual tag (and that associated with his/her company). His/her employees may also have provided ratings and/or annotations, which are further descriptive of his/her personality, leadership/management style and skills, work environment which s/he promotes and overall quality. A previous employee may allow him/herself to be contacted by the prospective candidate (e.g., in exchange for a fee).

Access Privileges Information - Users in an organization are frequently given privileged access to certain files within a corporate intranet but not others. Though there are many ways of profiling users according to their level of access privileges to information, the following example is considered: Based upon the position (e.g., responsibilities and tenure with the organization), users may be “classified” into groups according to different levels of access to confidential information. Virtual tags may be used to extend the capability by providing for immediate disclosure of a user’s information access privileges to another employee in real-time and in a physical context. Also, if a user reads or accesses certain information, meets with a certain colleague or friend, then a user might send a message X. This message could be (for e.g.) a request to perform some task relating to part of that information, a reminder to address certain issue(s) while chatting with the colleague etc. or, per the request of an individual’s employer or colleague if a given individual (a sales person) meets with user X send him/her message Y (which may refer to a previous encounter, experience or fact s/he should know pertaining to user X and which may have bearing upon their conversation or professional interaction).

Access restricted physical areas. In this case the virtual tag effectively may behave like an “electronic door key”. A variation of the technique may be used for

granting access to professional meetings, where information access privileges of users must match the anticipated confidentiality parameters for the scheduled meeting. Another application may include the ability to automatically enable access or restrict access, based on payment of fees, and whether or not an individual is a representative or partner of a competing company.

2.2.3 Business Introduction System

Another application domain for privacy-protected match-making, where users are anonymous until an agreement is struck is business to business introductions. For example, it might be useful to automatically identify synergies between businesses (e.g. in infrastructure, technology, or product) -- for the purposes of pursuing an advantageous strategic relationship. If the meeting is between two employees of competing companies, then the system of match making could also ensure that a meeting is predicated on a particular task that does not cause conflicts with their respective companies.

Each business might state within SDI: 1) what entities or types of entities s/he would share resources with. 2) If so, on a per-entity or per-entity type basis, what types of their resources (e.g. type of skilled employee and for what TYPE of out-sourced task) would the entity share. It is an obvious extension to look at sharing of code, technology, intellectual property. A major challenge and limiting factor being how well informed SDI, the neutral intermediary can be made aware of the needs/requirements of a company such that it can make evaluations entirely on its own regarding highly confidential materials with which it can accurately predict the basis for a deal WITHOUT disclosing to the prospective recipient what the technology or know-how entails (which could compromise the value of that asset should a deal not eventuate).

Corporations may also use SDI to utilize the above information regarding their human and technology sharing synergies in order to detect and recommend strategic (e.g. equity sharing, merger, acquisition etc.) relationship opportunities between the entities. B to b and even b to c user centric SDI user behavior analysis combined with text analysis should also provide revealing clues about what types of companies tend to share similar customers and provide similar (complementary or competitive) products and services which may suggest that such synergies are potentially available. Of course the vendor centric SDI service may

benefit from this information as well as provide information for other vendor identification of competition. Other data resources including electric payment protocols, EDI, automatic check payment, check services, etc. may be useful data resources as well.

Again the disclosure of detailed business information is very helpful and a data release policy defining the parameters for such strategic initiatives may be critical in order to determine what companies may be potential candidates for which initial feelers (of high level information disclosure) would be appropriate to put out to a prospective company to determine mutual interest and/or further basis for expected synergies.

2.2.4 Dynamic Annotation/Information Filtering

In this extended application of SDI, we allow users and other third parties to annotate objects (physical and virtual) with meta-information, either to remind themselves about a previous interaction in the future—or as a system of “knowledge learning”, where systems of users leave useful information for other users. Information is left in the environment, leaving a trail for other users.

For example, the information that is tagged to an object, referred to as a “virtual tag”, can contain a pointer to other relevant information, such as a survey of a film by a third party, or the user’s own comments/feedback. For example, a restaurant listing could be annotated with meta-information about the quality of the food and service. Such information, when provided by a wide sample of users, can provide robust information about objects. The information that is used by a particular user can be filtered—for example, weighting the opinion of a respected restaurant critic, or weighting the opinion of users with common profiles (when that information is available).

Virtual tags (i.e. meta-information) can be assigned to objects with physical locations, and the information triggered based on the physical location of a user (using LEIA technology). Virtual tags can be assigned with expiration dates or other time-sensitive information. An individual user might leave an “action item”, for example—next time I return to this object (e.g. web page/ vendor) be sure to perform this task, enter this query, check this link for new information. As another example, after a conversation with an SDI-enabled user it is possible to tag that user with some notes, to remember the conversation the next time the two users meet.

1 The technical innovation that allows this use of virtual tags, in addition to the protection of
2 privacy, is that we allow users to annotate information to objects that they do not directly
3 own through a system that separates virtual tags from the content that is tagged. In
4 particular, tags can be stored (either at the ISP-level proxy, or main SDI server) for
5 associated web pages, and exchanged/retrieved automatically when the object is accessed.
6 The virtual tags can be used in conjunction with target-object profiles that are generated
7 through SDI for web pages (and approved by vendors). Virtual tags can be searched, using
8 relevant terms, locations, or times, and can also contain links to authoritative information,
9 such as audio and/or video.

10 For example, in accordance with the prediction that readily deployable visual video
11 recording devices will become commonplace (even integrated into wearable computing
12 hardware). Users who are amenable to releasing such information under terms of their
13 privacy policy may allow other individuals subsequently visiting the same physical location
14 physically or virtually to gain access to such information. Such information may range from
15 detailed accounts, assessments of value, etc. A user could, for example, do a general query
16 about locations e.g., what is the percentage breakdown of types by their pseudonymous
17 attributes, who frequently visit this site which sites tend to have commentary of a particular
18 topical nature of interest to user U, which sites are visited frequently by individuals similar
19 to user U, which sites have had an event of a particular type or one which is similar to user
20 U's user profile occur in the recent past (or where or how close did such event occur with
21 respect to the recent location of user U and so forth).

22 Tags are encrypted, so that only SDI-enabled users can access them. Tags are also
23 associated with the pseudonymous ID of the user that left the information (although they can
24 be anonymous, an associated profile allows more accurate collaborative filtering
25 techniques). Finally, users can leave data-disclosure policies, embedded into tags—to certify
26 the properties of other users necessary to release the information. When tags automatically
27 are time-stamped with location, and time, and other information we allow for this
28 information to be “fuzzed”, as disclosed in the section on Randomized Aggregates, to
29 protect a user's identity.

30 In the physical world, implementation of meta-information in a user's physical information,
31 can be viewed via head-up displays, video cam monitors, wearable computing devices, or

1 The general application framework of matching users based on mutually beneficial and
2 acceptable terms and conditions within a physical context, e.g., be it e commerce (as above
3 described) business or social objectives, it is also useful to develop a mode which considers
4 the schedules of the calendaring agents of the individuals, as well a. s the timing and (if
5 relevant) the meeting opportunities which would be based upon mutually compatible
6 meeting venues which are mutually time compatible). A statistical model could be
7 developed which would take the above data (where appropriate) the geographic proximities
8 (considering lengths of travel routes between users may travel to attend a meeting) in order
9 to determine:

10 a. How to minimize the travel distance for individuals whose automatically scheduled
11 meetings are combined as part of an existing travel activity might be if the meeting was
12 scheduled, in part because it was of relevant values to the party (and it was at least relatively
13 close to the existing intended travel route).

14 c) It was sufficiently important to mandate a separate trip. The idea is to determine the
15 optimal physical locations of individuals to physically reside, (i.e. for residents or
16 even hotel lodging as well as possible the physical locations of business such that
17 when the above user agents – agent introduction meeting and scheduling
18 functionalities for physical meetings is performed for the amount of physical
19 distance which must be traversed is optimally optimized. The statistical method
20 used in this calculation would be similar to a scheduling optimization method used
21 to schedule flights and associated routes to minimize the total distance traveled by
22 all flight routes collectively (see Princeton publication using genetic computing
23 techniques). Of course, the above approach would be used for a more immediately
24 practical application in which the objective would be to simply reduce the amount of
25 distance a user traveled for his/her daily activities and to tie this framework into an
26 existing scheduling system which is appropriate for the user. It is possible to also
27 further define the ultimate objective to be reducing the actual travel time which a
28 user spends in a vehicle (thus the optimization algorithm would further factor in the
29 effect of time delays due to highway congestion which would affect the ultimate
30 ideal locations of these physical sites.

1 such as terms, involvement on a control and/or management level (as well as compatible
2 commercial technology venue in which they participate). The key objective is to identify as
3 first priority the commercial entities and customers which can provide the greatest degree of
4 benefit to the vendor then to, secondly, determine which entities within the VC/financing
5 community are able to bring the most valuable of these contacts to the table.

6 An investor may fund a small start-up and also improve the value of the start-up by
7 facilitating a larger vendor becoming a customer. An investor might negotiate a special deal
8 for such a vendor. In one variation SDI could be used to mediate the introduction of
9 optimally strategic investment opportunities to an investor *and* to identify strategic synergies
10 among and between different commercial entities who are already funded by an investor(s).
11 In this way it may be possible to introduce these investors and create joint participation with
12 the new prospective investor.

13 By leveraging SDI, the investors may even identify certain intriguing facts such as that the
14 synergies between one of their own customers and the other investor are also significant
15 and/or even dis-synergies exist between their own investors and themselves which do not
16 exist with other investors (e.g. and investment in a competitor or customer of a competitor)
17 that it would be mutually advantageous for the investors to entrust SDI with the job of
18 making appropriate introductions for re-distributing equity ownership of the investors into
19 those other companies which are more appropriate from a strategic standpoint.

20 The motivation behind agreeing to these introductions of course must be bona fide, serious
21 and not initiated with the malicious intent to capture otherwise confidential information
22 (particularly for those investors at higher level financing levels such as higher level VCs,
23 investment bankers and more generally investors who tend to be more minimally involved
24 in direct management or marketing initiatives for the company) it is certainly reasonable to
25 imagine that some of these investors may wish to apply some of the above suggested
26 techniques for finding (or even exchange equity opportunities within the commercial market
27 partially for the purpose of diversifying their risk by thus owning smaller equity stakes in
28 more companies and particularly those which are most commercially strategic and mutually
29 synergistic in nature.

30 The powerful data collections of SDI regarding comprehensive commercial/investment data
31 which of relevance to predicting the nature of an market dynamics of the associated

1 appropriate individual experts to evaluate that idea, which is either within SDI or within the
 2 organization which most likely needs the idea. This routing scheme could be based partially
 3 upon attributes associated with experts and features extracted from the description of an
 4 idea.

5 **2.2.8 Negotiation Intermediation**

6
 7 Given the above general application framework for SDI in which commercial entities can
 8 securely entrust to SDI with their commercial marketing and technology related focus and
 9 objectives, it is a reasonable extension to further extend this data Rich framework to enable
 10 the above described exchange of confidential information among and between commercial
 11 entities to occur not merely in a purely general and non-dynamic fashion but to additionally
 12 incorporate timely information which may represent time sensitive and critical decision
 13 processes which are in the process of occurring at that particular time and which may be
 14 relevant to the mutual strategic discussions and negotiations between the entities. A very
 15 important application/example of this approach is the fact that commercial entities may be in
 16 the process of negotiation with another commercial entity regarding particular commercial
 17 or technology related opportunity very often (the majority of instances) the nature of these
 18 discussions and particularly the identity (identities) is maintained as strictly confidential
 19 information throughout the discourse of these negotiations. Though the following solution
 20 may not be appropriate for all cases of confidential discussions and negotiations there may
 21 be particular instances in which a knowledgeable "expert" within SDI could be
 22 confidentially kept apprised of the details of the negotiations as well as progress and any
 23 obstacles as they occur. Of course, SDI and the expert have additional knowledge about the
 24 general and temporally specific needs and objectives of other third parties who may be
 25 potentially interested in the nature of the deal in progress.

26 At a general level there may be the opportunity for the expert to provide high level probing
 27 questions to the appropriate parties or decision makers within the other commercial entities
 28 which certainly do not reveal any identifying information about the negotiating parties, do
 29 not reveal any proprietary technical details which would compromise the proprietary nature
 30 of this information and enable the identity(identities)of the parties to be indirectly inferred in
 31 this regard. However, business terms and perceived value exchange may be thus conveyed

1 employed in this case in which it is possible to reward very well qualified (and truthful)
2 opinions on a very broad range of issues and domains in which the accuracy of the
3 ultimate outcome of a particular opinion is directly rewarded. This mechanism effectively
4 insures the quality of the referrals/recommendations and is considerably more efficient
5 then immediate payments for referrals where quality and truthfulness of the expert advice
6 cannot be audited prior to payment, thus creating a disincentive to provide truly
7 “valuable” advice.

8 The system of SDI can *forward requests for information to experts within the same system*
9 *as time-of-purchase competition, SDI acts as a CLEARING HOUSE for requests for*
10 *information, and experts can bid to provide responses.* The system of SDI can also help
11 users to choose between experts, based on proficiency profiles that are derived from the
12 value of information that experts have provided to other users. We can allow experts to bid
13 for the right to annotate and provide ratings for particular pieces of information, products,
14 services, etc.

15 In a web annotation example, as users provide annotations and recommend links, and other
16 users provide feedback about annotations, then the system of SDI can build a “proficiency
17 profile” for a user, to indicate the ability of a user to add value to the browsing experiences
18 of other users. The information retrieval and document clustering methods as taught within
19 the parent case #57549398 provide a statistically sound methodology to develop a user
20 profile that predicts the “proficiency” of a user to provide recommendations about objects
21 that fall in particular clusters.

22 In the economic variation we can allow agents to bid for the right to receive high quality
23 recommendations, at least in the case where recommendations are new and cannot be
24 duplication on the basis of current recommendations. We can create an “experts market”,
25 where experts are assessed on their ability to provide quality to previous agents, as noted by
26 those agents. SDI could also suggest a list of content categories that an expert might
27 consider operating within. We can also apply the rapid profiling techniques in the parent
28 case to assess a user’s expertise based on his/her response to a set of questions. A rapid
29 profiling tree is essentially a decision tree, which can be used to present a sequence of items
30 to which a user “responds”. The rapid profiling method enables a user profile to be
31 constructed in the shortest possible sequence of inquiries to the user. In this case, the

1 response is the user's recommendation links (which are then judged). With the benefit of a
2 comprehensive proficiency profile of the user the system may identify and present items to
3 the user automatically (if s/he is amenable) for which the user is predicted to be proficient.

4 We may pay individuals based on bonus points or coupons with future redeemable value
5 across a network of vendor sites (e.g., within that particular SDI recommended alliance of
6 vendor sites which may be dynamically generated for each individual user). Experts that
7 include links to pages could be paid for the value of increased value based on those links. Of
8 course SDI can continue to ensure that users are not motivated to provide biased links
9 because of this type of financial incentive, continuing to monitor how useful other users find
10 the links and recommendations.

11 The market for experts and real-time information filtering can be applied to new content.
12 Via collaborative-filtering based systems we can push new articles to appropriate users, so
13 long as experts provide initial scores for articles. In such an automated news system users
14 that represent particular types of customers can be paid to provide ratings of incoming news
15 articles. Based on profile information these ratings then imply ratings for the customers of a
16 news cutting service, therefore allowing filtering of news to the most appropriate and
17 narrowly defined subgroup of users of interests.

18 It can further be combined within the framework of the present incentive scheme to
19 encourage users who tend to individually find news, which is of particular relevance to their
20 daily lives and unique interests to proactively collect information (including but not limited
21 to digital recorded audio/video) and provide that information to news distribution systems.

22 Collaborative filtering leveraging both the overall quality/interest of user content as well as
23 of the nature of the new content which that particular user is providing can provide a nice
24 method to filter and channel this information.

25 The above bidding scheme enables a highly individualized incentive driven distributed news
26 reporting system, which can significantly increase the availability of content for individuals.

27 Users can continue to rate specific stories that they receive, so that the profiles of users are
28 dynamically updated, and experts can be swapped in and out based on the assessed value
29 and accuracy of new clipping.

30 The present clustering techniques may also be used to identify users that are able to provide
31 useful new content. This content may range from commentary, opinions, critiques and

comparisons, but can typically be more robust than simple annotations, and may include articles by knowledgeable experts providing pertinent or new information which is interesting or timely within the present context of the site. The predicted value to the particular cluster of users is a function of both the object profile of the original page as well as (perhaps even more relevantly) the object profile of the new content.

We can also have experts bid for the right to participate in online discussions, where experts are selected that are polarized on a current controversial issue, in order to fiercely debate a particular issue. We could also select representative group members, from within different primary clusters, based on the domain knowledge, skills, life experiences, preferences social/political or religious view or opinions of users, to engage in an active dialogue, interview or debate with the high profile individuals. Other members of clusters can be invited to observe a discussion, and to provide feedback to maintain information about the quality of experts.

2.3.1 Expert Sales Consultants

A static application of the information in the system of SDI to on-line sales is to use the information to generate an ideal list of prospects for sales agents, and then help sales agents to target products and offers to individuals. We can provide information to salespeople about users, according to the profile of a salesperson (and reputation), and a user's personal terms for data-disclosure. Similarly, a system of SDI in conjunction with the methods taught in co-pending patent allow automatic detection of salespeople close to users (via an anonymous location market). The market allows matches to be made, but does not reveal anything about a user that the user does not authorize. User profiles can also be used to determine responses to offers and products (see the methods in patent application "System for the Automatic Determination of Customized Prices and Promotions" [INSERT PATENT OFFICE NO]). SDI allows profiles to be built from extended interactions across multiple vendors, so long as the user authorizes the same pseudonym for each vendor.

As an example, consider an SDI-based system for contacting prospective customers and making personalized offers. The first step is to access the SDI pseudonymous user data base and identify the ideal audience of prospects for a product offering(s); then the second-step is to use the pseudonymous request protocols to request a sales call dialogue (or some other

1 form of dialogue) with the target prospect. The user's agent will then typically request
2 further information (either before or after allowing access to even the pseudonymous profile
3 database) such as what is the type of product (even possibly further associated details), and
4 an associated approximate price range (which may be either negotiable or non-negotiable).
5 The user may agree to allow a communication in the form of a targeted email that contains
6 information content, or a link to a web page with additional information. This Web link
7 and/or the e-mail can be *personalized* for the user, based on the user's profile. An email link
8 on the web page could be used for direct Web-based IP telephony connection to the user
9 (seller). Alternatively, a user's initial contact to web site could include encoded information
10 about its profile, where the vendor can only receive the profile information if it can
11 demonstrate certain profile properties. Large customers might then also receive specially
12 hand-crafted information, if it is indicated that there is a large potential contract. The user is
13 provided with a unique URL, or preferably the site is initialized so that when its pseudonym
14 is provided the offer is customized appropriately.

15 A more interesting example is a dynamic "experts market", which is an application of
16 dynamic matchmaking between users in real-time based on their profile information is to a
17 system for a *virtual sales force*. Users can also use seller profiles, to decide whether or not to
18 interact with a seller. The profile of a sales-person may show correlations between product
19 sell-rate and the type of product, type of user, that the sales-person interacts with. Initially
20 seller profiles may not be very well related to sales-performance, but instead based on
21 general SDI-style profiling, and wider (e.g. professional) credentials. Later, as a seller gains
22 experience, profiling can be based on a sales-person's track record (and this will subsume
23 other information).

24 A consumer can indicate its preferences in its profile for "approach" by a sales agent, and
25 then a match can be made with an appropriate sales agent dynamically in real-time via the
26 above agent-agent matching and introduction process. The sales agent may be compensated
27 when introductions lead to sales. The user could also receive preferential terms for a
28 purchase, that can be realized online or in a physical store via coded print-coupons. This
29 code with the time and identity of the (prospective) buyer as disclosed by the sales person.
30 The buyer must authenticate that sales interaction (e.g., via disclosure of the code).

1 The criteria for matching the seller with a qualified prospective buyer is represented by
2 client-provided meta-information associated with profile information, and can be used
3 within a collaborative filtering system to determine the level of predicted interest which the
4 buyer is likely to have for the items offered by a particular seller. The product space
5 predicted to be useful to a buyer can be predicted based on his/her profile, browsing
6 patterns, etc. Context can also play an important role, for example is the buyer currently
7 engaged in browsing or buying related activities, can the buyer be persuaded to buy now
8 with a good offer, or will the buyer never buy until he/she has performed more product
9 search?

10 Vendors may also compete for sellers based on the profile information and success-profiles
11 of a seller, and a perceived match with the products offered by the vendor. A seller may also
12 work for numerous categories of products/services and vendors. Finally, the sellers in the
13 market could enter into a market-based system, so that buyers bid for the right to work with
14 a particular sales person.

15 Two other aspects of the present access to or by a prospective sales person include

16 (1) In an off-line context, we can use a LEIA-based method to identify buyers and sellers
17 with similar profiles in physical space, and dynamically reroute their paths to allow a
18 meeting to discuss a possible trade.

19 (2) Personal "Chemistry". In addition to a pure "product-space" set of profile features, the
20 system might also consider wider compatibility between sellers and buyers in making
21 meetings, for example choosing to introduce agents that share similar hobbies and spare-
22 time activities.

23 The problem might also be informational: e.g. find an expert on ancient American
24 civilization for purposes of writing an article, or answering a specific question. Relevant
25 information might include the expert's resume, and the expert's knowledge expertise profile
26 developed from his/her activities in responding to previous queries.

27 We might use a "fuzzy rule" to determine whether a user has a profile that is sufficiently
28 close to an agent's goal profile to allow an agent-agent interaction and exchange of
29 information. When the rule does not quite fire the system of SDI might also seek to clarify
30 points of uncertainty, requesting further information until there is enough information to
31 decide on the appropriateness of a contact.

1 profile of documents or web pages being interacted with. Activities may further be inferred
2 by such indicators such as the particular physical location of the user (i.e. within a LEIA-
3 based system [INSERT US PATENT NO]): the particular sequence of movements, the
4 particular types of devices with which the user is interacting, etc.

5 More complex rules are certainly also conceivable, such, as the user will not interact with
6 other users except under specific conditions, i.e. I will only talk to a user of type X
7 (salesperson) if I have done nothing for the past 10 minutes and have nothing scheduled in
8 my meeting planner. Such rules are initially hard-coded, but can then be learned and/or
9 validated based on inductive learning: i.e. is it the case that I am more receptive to certain
10 types of activities at particular times of the day, etc?

11 In some cases SDI may identify individuals whose personal data revelation requires certain
12 minimal requested personal data from the other party as a condition to an interaction, and
13 SDI can again be used to automate this process is if it is well explained in a user's profile.
14 When confidence is low the user may request additional information (anonymously, via
15 SDI) and make a personal judgment in real time as to whether to continue an interaction and
16 reveal more information, and whether or not an introduction is desirable.

17 When information is valuable and sensitive, then it is important that initial information
18 exchange is via SDI, without the end-users receiving explicit information. In overview, a
19 user states the types of conditions that are necessary for a continued one-on-one interaction,
20 and the system of SDI states whether or not conditions have been met without giving any
21 specifics. Of course the identities of the agents are not released until introductions are made.
22 SDI can track all activities, behaviors, and correspondences in which each user engages, and
23 build a complete picture of the profile of a user with regards to interactions and the
24 exchange of information. SDI may also use *human* experts to interpret requests and build
25 rule-based systems to semi-automate decisions and automate easy cases.

26 A simple but related application could be applied in the e-commerce domain to
27 connect on-line customers with human sales representatives. A current version of this,
28 offered by eStara, uses Internet-based telephony technology to enable browsers of on-line
29 catalog pages to talk with knowledgeable sales representatives, simply by clicking on a
30 "talk" button displayed on the web page of interest. Integrated into this methodology is
31 the ability of the sales representative to identify which particular web page the visitor is

1 currently viewing, thus enabling him/her to more efficiently and intelligently assess the
2 needs of the customer.

3
4 This methodology could be further enhanced through the tracking of the user's
5 pseudonym and associated profile data provided via SDI, which represents a far richer set
6 of information than that used by eStara (which only makes use of the current web page
7 identifier). A customer's full profile can include demographics, previous purchases,
8 previous web-site visits, physical measurements (for clothing purchases), health history,
9 and income.

10
11 In a sense, individual sales representatives also have personal profiles; these can
12 include experience, demographics, languages spoken, previous customer contacts, and
13 product knowledge. An especially important factor is a representative's relative success
14 in selling particular products – this history may reflect various personal strengths and
15 weaknesses.

16
17 When the customer clicks on an on-line catalog's "talk" button, his profile is
18 immediately forwarded to the sales system, along with the particular product he is
19 viewing. Matching algorithms (as described elsewhere in the patent) then choose the
20 most relevant sales representative currently available for on-line communications. The
21 representative is presented with a screen containing facts pertinent to the sale (including
22 product details and extracts of the customer's profile), and he is then linked to the
23 customer via Internet telephony.

24
25 Extensions to this technology include:

26
27 1. Customized Price Discounts and Promotional Offer Recommendations

28
29 By combining the predicted affinity of a user toward a particular item with price
30 elasticity metrics (gleaned from his/her transaction logs and from real-time feedback
31 provided to the system by the sales representative (e.g., the customer appears to be in a

happy and spending mood)), it is possible to calculate a “personalized” price that optimizes the expected profit from a sale. This optimized sales price is communicated to the sales representative, who is authorized to offer the discount to the customer. Digital coupon technology would be used to ensure that the offer is redeemed by that customer exclusively, and within a pre-set time limit.

2. Automated Sales Representative Support

While interacting with the customer, the sales representative is presented with constantly-updated screen that provides information to support his sales effort.

a. Rapid Profiling

If not much information is known about a customer, the representative would be guided through a script designed to elicit the maximum amount of information in the least number of questions. This information would be used to rapidly supplement the customer’s profile.

b. Customized Sales Scripts

The representative’s scripted sales pitch would be dynamically adjusted to reflect the profile and current attitude of the customer. For example, if a customer is having trouble making a decision to buy (which could be detected by voice analysis tools and the total time elapsed in the transaction), especially persuasive text, geared to the customer’s demographic, would be presented to the representative.

c. Additional Offers

If separate, but related, products are predicted to be of special interest to the customer, the sales representative is alerted so that he can promote the additional offers while still in contact with the customer.

Barter Exchange for Personal Favors

In this variation, SDI can act as an exchange where professionals can exchange personal favors, for example one agent might indicate that it wishes to be introduced to another agent that can help with a particular professional problem. Both buyers and sellers can submit their respective needs, for example the need for a particular personal or professional favor, and in the case of sellers the ability to provide favors. Other personal information, for example: professional, political contacts, organizational affiliations, areas of competence, professional responsibility and spheres of influence etc. can be confidentially entrusted with SDI and used to make particular types of matches.

SDI can initially automatically define appropriate matches between the “buyers” and “sellers”. Given sufficient data, humans may use content analysis techniques to define rules based upon certain key examples or commonly occurring request and matching solution pairs. The techniques of the parent patent application US Patent # 5754938 can also be used to refine rules, and create new rules, for the cases in which the level of confidence with introductions based with the current rule set is too low. The mechanism for establishing the value at which favors are sold might involve a bidding scheme, but in this case professionals may bid a “personal favor” in return for another favor. I.e. the transaction becomes disintermediated from the outside economy, with a pure non-money based method for exchanging favors. A user with the ability to provide a favor receives bids from other users in need of the favor she/he is able to provide. Of course, the value of a given favor is likely to be worth more to some individuals than others.

The nature of the favor may range from very small and mundane, e.g., providing professional or personal advice or answers to a question, or it may be very significant and resource intensive from the standpoint of the provider, e.g., making a professional recommendation for a job position or political decision or providing recommendations/referrals for clients on a professional level or providing business recommendations to a partner or affiliate regarding a given individual or commercial entity with which to do business.

The techniques of peer-ratings within a reputation system can be introduced, and extended to cover a wide array of qualitative descriptive attributes to reveal the qualifications and ability of an individual, to allow some individuals in the bidding process to be excluded in

an initial “filtering process”, because they fall below expected standards of service and performance. An auction based model may establish a price for a particular type of favor, and for a particular individual, and therefore users with high performance ratings could be expected to sell for higher prices than other agents.

BARTER EXCHANGE FOR FAVORS

The qualification threshold of the provider of the favor is more important in some applications than in others. For example, if the favor involves an interaction with a particular individual for organization (e.g. a political favor) in which the individual’s anticipated ability and qualifications may be a significant factor in the likelihood to achieve the desired results. And thresholds may be set relating to quality or qualifications in their regard depending upon the relevance of this factor and/or the importance of the favor, which depends upon the type of favor requested. The bidding process is typically a reverse auction in which the request favor is passed around to multiple individuals (or entities) which SDI deems qualified for the particular favor requested (thus the lowest price which fits the qualifications criteria is typically the offer which is accepted. In other (perhaps many) of the barter exchange for favors, the qualifications of the requester may be an important factor for the requestee. These qualifications may include, not only professional abilities, proficiencies and credentials, but potentially much more subtle attributes relating to the individual’s interpersonal, social or psychological profile and/or behavioral profile, for example, how SDI predicts that the individual receiving the favor will integrate and adapt within a particular professional context and/or interpersonal context which defines the situation and/or environment with which that individual must interact and/or perform, e.g., as in a business deal, a professional opportunity, an educational opportunity, social opportunity, etc. In order for this pre-qualification of the requester to unilaterally work so as to assure the requestee with the level of confidence and trust in the requester which approaches that of a traditional favor (in which the requester and requestee are typically extremely knowledgeable and trusting in one another) requires a very efficient and perhaps fairly comprehensive reputation system revealing much about the above described personal and professional aspects of one of both parties. As suggested, in such a system and just as

1 important, is the trustworthiness of the individual providing his/her opinion about the
2 credentialed profiles of the parties. Indeed a political or decision maker in a large
3 organization would be extremely remiss (and perhaps even politically harmed) if an office
4 or position were provided to another individual (requester) in which weaknesses in that
5 individual's professional and/or personality later became evident to the detriment of the
6 organization or political entity, as an example.

7
8 It is even possible, that in some cases, the need may exist for a pricing model in which the
9 individual providing these credentials through the reputation systems is compensated for the
10 task. If such an individual must be fairly knowledgeable and assure to properly judge
11 professional or subtle personality components, this individual may develop an independent
12 reputation for effectively and truthfully judging these certain characteristics on a category or
13 domain basis. If the subset of "judges" is fully distributed and ubiquitous permeating a
14 substantial portion of the population as a default, if the qualifications of the judge are unable
15 to be validated in advance, it may be preferable to provide payment following a reasonable
16 level competency and success on the part of the requestee and requester. Or, in another
17 variation, users are required whenever appreciated (and in accordance with their own
18 judging qualifications) to provide opinions on the individuals (or other entities), and they
19 must comply up to a certain quota and with truthful and accurate opinions in order to
20 maintain privileges to engage in the service (e.g., www.favor.com) or other applications or
21 services resulting from the reputation system. It is also, of course, critical to ensure the
22 judge possesses with regards to the individual or entity with which he/she is affiliated. In
23 this case, the assumption is that the barter exchange for favors could cover nearly any type
24 of favor desired by a user and the reputation system is implemented for most individuals.
25 Of course, the present system may also be extended within a business context in which the
26 favors include those provided by in between businesses. In both the individual and business
27 application context, it is necessary for the above described use of privacy policies to dictate
28 what types of individuals or entities a user or organization to include, exclude or price
29 discriminations against e.g. to various degrees).

30

1 The following section "Agent-Mediated Value Exchange in the Supply Chain" talks about
2 an economic system methodology in which a "value chain" is established for which entities
3 may compensate other entities for present or often future value, which is likely to be
4 sustained as a result of certain actions performed on behalf of that entity. This value often
5 conveyed in the form of "barterables" could also be provided in the form of "favors"
6 between the commercial entities (or potentially in a variation, individuals).

7 8 **2.3.2 A Market for Referrals**

9
10 Situations frequently arise in a variety of contexts of human interaction (whether social or
11 professional) in which a user may wish to refer the user they are in contact with to
12 another individual. Often this occurs in a professional services context which a user has a
13 particular need/or other characteristics which make him/her an appropriate match for the
14 services provided by the other party. Or in a business context, often a user will forward a
15 business contact or associate to another colleague who is deemed more appropriate for
16 the particular context and/or scope of business. Likewise, in a personal or social context
17 users may sometimes meet two or more individuals which they observe or perceive share
18 common interests, goals or beliefs or perhaps possess complementary capabilities,
19 knowledge, or characteristics. In each of the above scenarios, virtual tags may provide
20 substantial benefits. For example, the referring user could forward the relevant portion of
21 the profile and identified need of the user to the referring party whose user agent may
22 determine the acceptability of the request and/or the priority with which a communication
23 or meeting could be scheduled (e.g., as could be automatically arranged by/between the
24 two party' calendaring agents). If the referring party's agent is unable to make a decision
25 or priority assessment for scheduling purposes) on behalf of the user, the agent could
26 instead try to contact the individual him/herself for assistance (and statistical feedback to
27 the system's data model). In order for these types of referrals to be performed efficiently,
28 the area of expertise required can be specified, and provisions can be made about the type
29 of referrals that a professional will accept.

enable free access by users on the Internet who may want to, and are able to, usefully contribute. The idea behind the present scheme is to leverage techniques of natural language processing and/or information retrieval in order to develop profiles of individuals based upon the value of their past contributions to documents; for example informational content, style of authoring, etc. In a Web-based application the content can be graphical designed content, perhaps even video segments. Within SDI it is possible, based on feedback from previous articles, to anticipate how valuable the information is likely to be for a particular readership.

The problem might also be informational: e.g. find an expert on ancient American civilization for purposes of writing an article, or answering a specific question. Relevant information might include the expert's resume, and the expert's knowledge expertise profile developed from his/her activities in responding to previous queries. Level of expertise might also include the size of projects performed within a particular specialized area, and relevant education qualifications.

2.3.4 A System for Smart Consumer Research

There is a significant need within the field of consumer product creation and development to be able to more instantaneously, and on a larger scale, collect direct feedback from a large number of consumers in lieu of much more sparsely populated focus groups and "product development" experts. There is also a need to better characterize consumers whose feedback is utilized in any kind of market test analysis. Collaborative filtering can provide a very efficient solution to both of these problems by enabling focus of feedback from selectively chosen consumers whose attitudes and opinions (and even marketing ideas) are the most meaningful and representative of the large segments, comprising the majority of the consumer population for that product. We can further extend this methodology to determine and measure the "value" of certain individuals in the process of product assessment, i.e. to identify the "experts" in a particular area.

The information in SDI can be used to identify appropriate clusters of users for a particular product, so that a prototype or actual product can be presented to users, and feedback collected. The goal is to collect detailed feedback: across a variety of quality and consumer

1 Another variation is an auction scenario, where a seller brings an item to market, and buyers
2 compete for the item. A professional, that makes a living by buying items and selling them
3 on, might like an exclusive SDI-based auction site. Within the system of SDI we can support
4 a *network of auction-sites*, that push agents around between sites and revenue-share, with
5 source web-sites providing shares of revenue achieved in subsequent purchases on other
6 sites. In another variation, the operator of a small site may be linked from a larger site, and
7 provide a share of revenues to that larger site. SDI has a couple of important roles:
8 monitoring purchases and ensuring that contracts are fulfilled, and also estimating the value
9 of links if an up-front price is to be negotiated, based on information about the profiles of
10 streams of customers to a particular site.

11 The information in SDI can be used to allow disintermediation of consumer-to-consumer
12 markets, with agents able to reach ideal prospects based on personal information stored in
13 the shared database, make an anonymous offer, and the consummate the trade of possible.
14 This is likely the way of the future in this consumer-to-consumer market. Just as a vendor in
15 current e-commerce systems can store information about its own user-base, and use that
16 information to send personalized offers to users, within the system of shard information in
17 SDI all agents can share information with all other agents, and the playing field is leveled.
18 Users can query the pseudonymous user profile data and (if permitted), gain direct personal
19 access to appropriate target customer prospect.

20 Conversely, prospective customers who are interested in a particular item(s) may query and
21 access the database of users who are knowledgeable regarding that particular
22 product/service (which may possibly be under a pseudonym), and also accessible to the user,
23 i.e. available at an appropriate time and in an appropriate location.

24 2.6 Transportation Example

25

26 Another application of user profile database, in the case that the database contains also real-
27 time information, is in an application to transport scheduling problems. The traffic systems
28 in many large systems are congested. Via the system of SDI it would be possible to allow
29 agents to represent the wishes of their uses within an SDI-enabled mechanism that
30 controlled access to certain lanes on highways. The goal might be to control the flow of
31 traffic, such that users agree to pool vehicles based on similarities across planned trips, and

1 users also payments in return for the right to travel at certain times and in certain lanes. The
2 goal is a more efficient system that maximizes the sum value to all agents.

3 This is an example of a more general application area, where agents can look for
4 opportunities to *change their actions and behavior in return for higher system efficiency,*
5 *and/or payments.* We describe below an application to user agents negotiating on behalf of
6 individuals ("travelers") to optimize the efficiency of a shared transportation system. The
7 fundamental idea is to embed multi-user awareness into the basic fabric of a transportation
8 system, in which the system aims to maximize the efficiency of the system by allowing
9 payments between individuals.

10 To accomplish this task, the system must be able to anticipate both the immediate direct
11 effects, a particular accommodative action will have on another agent(s) (mobile entities) as
12 well as the indirect (cumulative) effect on the other agents and the associated time delay
13 impact on each agent and on the entire transportation system as a whole.

14

15 The techniques of the LEIA can be efficiently employed in tracking the present and
16 anticipated activities, location and movement patterns of individuals who are in the process
17 of traveling to a destination via simple or multiple modes of transportation, e.g., any
18 combination of the following, i.e., pedestrians, automobiles, taxi, train, and bus (including
19 public transportation). LEIA can be applied in both scheduling the flow of traffic for
20 transportation media which are not subject to fixed time schedules (cabs, autos, pedestrians)
21 as well as to dynamically improve overall efficiency of the movement patterns in order to
22 assure that users on a collective basis arrive at their desired destinations with minimal delay.

23 The system of SDI can implement a real-time market where agents make and receive
24 payments in return for changed actions from other users. For example, agents with high
25 priority may be prepared to pay more than other agents for the right to enter a highway at
26 any particular moment in time.

27 The market pressures at any point in time are between the marginal cost to an individual for
28 changing his/her plan, and the cumulative marginal benefit to other agents in the system.
29 Within incentive-compatible mechanisms, such as the regular price-based market place (e.g.
30 the Walrasian tatonnement model) with enough agents, then it is optimal for an agent to

1 reveal its true value for different outcomes to the marketplace, and the system of SDI in
2 “clearing” the market can maximize system efficiency.

3 The central contribution of the SDI system is that it acts as a trusted auctioneer, receiving
4 information from agents, and implementing an outcome based on that information without
5 releasing that information to other agents. All that agents see are requested actions, and
6 payments received or payments to make.

7 We could model a core sample of individuals on an active basis to determine implied
8 valuation functions for different types of users, based on feedback provided by those users
9 about how happy they are at any point in time. Within an incentive compatible system it is
10 optimal for an agent to have a truthful and complete representation of its valuation function.

11 The problem is to discover that valuation function. Any help from SDI, via data mining
12 techniques (i.e. similarities between my profile and the profile of another user that has rated
13 his/her happiness) can be beneficial in reducing the costs of participation in such a scheme.

14 Then, before taking a journey, the user could provide some explicit information regarding
15 the nature of the travel, for example the level of importance in minimizing travel time,
16 whether the trip is business or sightseeing, the time of day, etc. With this information the
17 system can assess the user’s valuation function, and then have the user participate within the
18 market, and secure a plan based on actions from all agents. The system might compute price
19 thresholds, that represent different things the user can achieve for different prices. The
20 system based upon its determination of situations/context can presents the price and time
21 and urgency inferred for a user prior; which can be reviewed and corrected by the user.

22 With this profile information, the auctioneer can now compute prices to clear the market and
23 maximize economic efficiency in the system; taking the following approach:

- 24 1. Armed with statistical information about the users of the transportation system e.g.
25 the highway for the next hour, compute static (fixed) prices for different options.
26 The goal is to maximize the efficiency in the market through a simple pricing
27 mechanism, that is computed based on information provided to the auctioneer in
28 advance.
- 29 2. Announce the prices, and allow users to use the system as they desire, with
30 appropriate payments made. Payments can be readily extracted from agents via
31 automatic toll systems.

1 Of course over time the performance of such a system could be optimized, as the auctioneer
2 (the SDI marketplace) learns about consumption patterns of the agents over time. One way
3 to provide incentives for agents to provide information to the system to allow up-front
4 computation of useful prices to support an efficient outcome would be provide *discounts* to
5 the agents that provide information. When such a system works well then roadway traffic
6 can be controlled, congestion controlled, via an automated price-based system. Additional
7 information made available from within SDI might allow adjustments to prices, based on
8 unexpected flows of traffic. Just because the prices are optimized once up-front, based on
9 projections, and announced to agents using the system; it can still be possible to adjust
10 prices during any particular period—so long as there is a method for that feedback to alter
11 the actions of agents, i.e. there must always be an alternative to make such a system work.
12 Consider a two-road system, one shorter and one longer, then the longer route would be
13 priced less than the shorter route, and the price differential adjusted based on current levels
14 of congestion.

15 Again, to clarify, SDI has a key role in this system. Agents are only prepared to reveal
16 valuation and trip data within a system that carefully protects their privacy, and controls that
17 information.

18 Via the optimization performed to compute appropriate prices the model takes into account
19 the effect on congestion that consumption decisions have, i.e. there is a relationship between
20 the value to an agent that selects option A and the number of agents that select option A.
21 The auctioneer needs a model of the transportation system itself to perform this type of
22 computation. Essentially, we compute the fixed point of a non-linear system of equations.
23 Stochastic optimization techniques are suitable for such a calculation.

24 As the system collects more data it can be less important to require explicit information
25 from agents within the system, unless it is believed that there is a special situation about to
26 occur (for example a Flyers game), which will have unusual effects on traffic patterns.

27 Co-pending patent application entitled LEIA-TR describes flexible tools and automatic
28 traffic ticketing which could be integrated in to the same transaction infrastructure. SDI
29 would provide a methodology to elicit information from agents, and dynamically set the
30 price of traffic tickets and parking meter tolls to maximize the performance of a City's
31 parking resources. For example, one goal of the transportation authority can be to optimize

1 the amount of revenues it can achieve, by charging more during “high-demand” travel
2 periods.

3 Extensions of the present scheme could include coalition discounts, where coalitions of
4 users can negotiate discounts based on group purchasing power (let us all use your system
5 for a 10% discount or we will use an alternative road system). The coalitions may be formed
6 automatically within SDI using shared profile information, allowing agents with non-
7 competing interests to “pool their buying power”. Finally, it is interesting to note that users
8 who share identical travel objectives can be encouraged to share vehicles, and/or public
9 transport can be provided as alternative means of transportation on a dynamic basis.

10 **3. Client-Side Data Mining Applications**

11 Finally, we describe applications of *client-side data mining*, where the private information
12 about an agent remains on its local client machine. Personalized of services and products is
13 performed by pushing methods to the client machine, where the methods compute based on
14 local profile information the most appropriate information for the user. This is very useful
15 because the user retains absolute control over his/her profile information, but can still
16 receive the benefits of personalization. This is an extension of iamworthit: users still provide
17 profile information to the central database, to allow central queries and to receive value for
18 that information, but never release identified information to a vendor.

19 **3.1 Client-Side Personalization**

20 Client-side personalization allows vendors to push personalization rules to client machines
21 that are special queries which use information store in a local client profile database to
22 provide a user with personalized information and/or services. The personalization rules use
23 locally stored profile information to compute an optimal product or service, or to allow a
24 vendor to configure a virtual shop front. The answer to a query is returned to a vendor, to
25 allow the vendor to push suitable commands to the client to enable appropriate displays to
26 be created on the device with which the user interacts with the client.

27 Client side personalization is useful within the system of SDI because:

- 28 a) It allows vendors to leverage profile information across multiple profiles that
29 belong to an individual without explicitly receiving the information.
- 30 b) It provides users with the convenience of using their true identities in payment
31 and specification of delivery addresses for goods, because vendors do not need to

1 which can be maintained at a user's client for safe personalization, and pushed to the central
2 SDI data warehouse for collaborative filtering type applications.

3 One draw back for a vendor is that it provides the agent computer with methods to target
4 information. For example, Amazon.com must provide the agent computer of a consumer
5 with a method to allow it to personalize future transactions. The method is valuable
6 information to Amazon. However, to protect this method from other possibly competing
7 vendors, e.g. www.barnesnoble.com, Amazon can use encrypted codes to represent the
8 results of a profiling query on the user-side profile database, and these encrypted codes can
9 even change periodically, or from user to user.

10 In the basic variation of personalization the vendor pushes a personalization query to the
11 agent, to be executed on the agent's computer in the client-side query-execution module on
12 the basis of the agent's single-user data warehouse. This is illustrated in Figure 18 above.
13 The vendor then receives the result of the personalization query, which might state for
14 example: "this user is very like my concept of what user Mary might want to buy", where
15 Mary is an aggregate user that represents data mining the vendor has performed.
16 Alternatively, the result might say something more direct like "show the user product X", or
17 make special offer "Y". Whatever information the vendor receives in response to the query,
18 it can then push commands to a user's agent computer to display graphics that represent its
19 desired response. For example, in an Internet application with a user that is browsing via an
20 Internet browser, the vendor could push HTTP commands for a personalized shop front for
21 that user based on the results of its personalization query.

22 We can describe a number of variations on this basic concept, that describe different method
23 to compute the best action to take given a vendor's personalization method and a user's
24 profile information:

25 14 Every-time Client-side personalization. In this variation the vendor pushes a
26 personalization method every time the user initiates a new interaction. The agent
27 computer executes the personalization query, and makes its response to the
28 vendor. The vendor can then send personalized content to the user.

29 15 Periodic Client-side personalization. In this variation the vendor maintains (either
30 vendor-side or user-side) the data that the last personalization was performed and
31 pushes a new personalization query to the user periodically, when it thinks that

1 appropriate to show to a user is made on the basis of a similarity comparison between the
2 user's profile and the profile of other users for which the vendor has information about how
3 they responded to certain objects. The straightforward method is to pass the profiles for an
4 entire user population to the agent's client machine. This is undesirable when there are many
5 users in the population.

6 A better solution is to pass only cluster centers to the client, where the cluster centers are
7 chosen to allow collaborative filtering to a sufficient degree of accuracy. For example, we
8 illustrate below an efficient method for collaborative filtering at the client, and our proposed
9 efficient implementation. This is illustrated in Figure 20.

10

11 The idea is to select a subset of users that represent the entire space quite well, which the
12 vendor has collected by performing data mining queries on the central SDI data warehouse,
13 and then submit just those data points to the client. In the Figure we plot the complete set of
14 user profiles in profile space, with each user profile associated with information about how
15 to target a product or service to a user with that profile. A typical method to perform
16 collaborative filtering with data of this kind is to find profiles in the neighborhood of a new
17 data point, illustrated with * in the above diagram, and then base the decision on what was
18 successful or unsuccessful for those users. Clustering replaces a set of similar users with one
19 "cluster point" that represents what information a typical user in that region should be
20 shown. This can be computed using standard cluster analysis techniques. The reduced space
21 of data points, shown on the right, is sent to the client, and the client performs collaborative
22 filtering over the cluster points to computer an approximate solution to the full
23 personalization query. Completing the example, the agent computer can determine which
24 "aggregate user" its profile is closest to, for example Mary or Yu-wen, and send this to the
25 vendor.

26 Notice that the vendor does not even have to provide the "y-values", or the personalization
27 methods which are associated with each user profile, it is sufficient for the agent to compute
28 the closest cluster center and pass that information back to the vendor. This is useful to
29 vendors because the information is not useful to anyone but themselves, because no-one else
30 can interpret what it means to be like aggregate "Mary".

1 An alternative is to provide generic information to the client each time, and allow the client
2 to display appropriate information automatically, without requesting information from the
3 vendor. This removes the final handshake between client and vendor, and might be
4 important if the user does not wish the vendor to receive any information about the result of
5 personalization, not even how it decides to use the data provided by the vendor. The vendor
6 only gets information in the case of a purchase.

7 A simple extension, permitted with the architecture as described, because the agent
8 computer implements the central SDI data warehouse query-execution and pricing modules
9 but with only profile information about the single user, is to allow the user to demand
10 payment in return for allowing the vendor to make a personalized sales pitch.

11 **3.1.1 Implementation Techniques**

12 The implementation of our system for privacy-protected personalization requires the
13 following functionality:

- 14 (1) A vendor must be able to push generic information to a user's client machine
- 15 (2) A vendor must be able to push methods to a user's client machine that allow that
16 client to process the generic information locally, using a user's profile information
17 as an input into a decision theoretic rule (that can be rule-based, collaborative
18 filtering based, etc.)
- 19 (3) A client machine must be able to process the generic information locally, together
20 with a user's profile, and decide what information to display to the user.

21
22 A reasonable implementation of safe personalization within the system of Secure Data
23 Interchange can leverage current World Wide Web standardization efforts that are based
24 around the Extensible Meta Language (XML) [B97; CKR 97; KR97a; KR97b], and
25 promoted by the World Wide Web consortium [W3C-XMLabc]. XML allows data to be
26 structured and accompanied with a Data Description File that defines the semantics of the
27 data, making it suitable for client-side processing by the receiving agent computer. The
28 method to process the information can be for example implemented with the Java
29 programming language, which will execute on a variety of different platforms via a Java
30 Virtual machine. A vendor can send formatted data via XML and a Java program to target
31 data to the user based on access to the user's profile information.

1 It is important that the SDI client-side proxy agent carefully controls the access that the Java
2 method gets to an agent's profile information, but this can be structured in just the same way
3 as with query-execution in the central SDI data warehouse. The method must maintain a
4 separation between profile information and the execution state of the program, and check
5 that all data access commands are consistent with a user's price rules (and data access rules).
6 Furthermore, it is important that the SDI client-side proxy agent executes the vendor's
7 personalization query does not communicate profile information to other agents, and
8 methods can be used to control the access of the query to communication ports, memory and
9 disk space.

10 Clearly, standardization of user's profile information is necessary, so that all vendors can
11 write methods that can access a user's profile. One solution that would support
12 implementations from multiple providers of profiling functionality (i.e. would not require
13 clients running standard SDI-profiling software) would be to provide "XML-data
14 wrappers", that convert non-standard formats into a standard XML data format, that is used
15 by all personalization methods. One language for representing user profiles has been
16 proposed by the W3C consortium, known as PIDL (personalization definition language)
17 [W3C-PIDL].

18 3.2 Real-Time Targeted Advertising

19 A useful application of SDI allows vendors to provide targeted advertisements to users
20 based on their browsing behavior and other profile information, but without actually
21 receiving explicit information about a user's profile. We propose an advertising-server,
22 which controls the network of adverts on behalf of vendors and users. A user's client
23 machine can receive a choice of adverts whenever it hits a web site in the ianworthit
24 advertising network, with local evaluation based on the local profile information at a client
25 to decide on an appropriate advert to display.

26 The system is designed to support dynamic personalized advertising, with vendor-side
27 competition for the right to present an advert to a buyer, and buyer side criteria for accepting
28 or refusing an advert. A technical solution is presented to allow vendors to select an
29 appropriate bidding-policy, to maximize revenue from advertising. We use client-side
30 profiling and advertising selection to support targeted advertising with absolute guarantees

1 on the privacy of a buyer. The ad-server sends a selection of adverts to a user, and the user's
2 client chooses to display the most appropriate advert.

3 With a limited capacity to display an advert to a user, it is necessary to control the adverts
4 that vendors display. We propose an auction-based system to compute prices dynamically
5 though competition between vendors. The auction is a second-price sealed-bid auction that
6 encourages advertisers to bid truthfully without "gaming" the outcome of the auction. The
7 advertising auction is implemented on a user's client machine.

8 In overview, the following steps are performed to support targeted advertising to users:

- 9 1. Buyer hits the web page of a vendor with an iamworthit advertising link.
 - 10 2. Vendor pushes content to the buyer (possibly personalized).
 - 11 3. Vendor contacts iamworthit ad server, requests an advert and pushes any profile
12 information about the user.
 - 13 4. The advertising server represents a number of companies that wish to advertise
14 with appropriate users, and determines a short list of adverts to offer to the user's
15 client based on profile information.
 - 16 5. The user's client receives the choice of adverts and accepts an advert based on the
17 values of bids and the likely suitability of an advert to a user.
 - 18 6. The client requests the advert from the ad server, and then displays the ad.
- 19 Payment is collected for viewing the advert, and finally the user's response to the advert can
20 be stored in the central SDI database for future tuning of advertising policies.

21 Figure 21 illustrates the system. The novelty provided by the system of SDI is that: (a)
22 advertisers can decide when to place adverts and how much to bid based on historical
23 information stored in the central SDI data warehouse; (b) the final advert is selected
24 dynamically at the client with local profile information stored about a user. The stored
25 profile information about users stored within Secure Data Interchange allows vendors and
26 client-side SDI proxies to estimate the probability that a user will respond to an
27 advertisement, given its profile and records about how other users with similar profiles have
28 previously responded.

29 Vendors, represented within the advertising-network server have access to some information
30 about a user, and can compute an expected value of placing an advert, based on the final
31 purchases made by other users that were shown adverts. Meanwhile, client-side machines

1 have a more complete picture of a user's profile, and can compute the expected value of
2 viewing an advert based on the advert and the price offered by vendors. In general, as the
3 acceptability of an advert increases vendors will wish to bid more, and users will accept the
4 advert for less money. Competition between vendors drives the bid price up, and allows
5 users to receive the value of information about their profile to advertisers.

6 **3.2.1 Example: Technique to compute the expected value of an** 7 **advert.**

8
9 A collaborative filtering system can be used to predict the probability that a user with profile
10 x will respond to an advert. The computation is based on historical information in the central
11 SDI data warehouse for similar users to the advert, and also on the basis of historical
12 information for similar users to similar advertisements if there is little information available
13 about the actual advertisement. Information can be provided to advertisers anonymously in
14 performing queries, and also randomized if necessary.

15 The basic technique is to select a set of users that have been shown the advert, or a similar
16 advert, and are close in profile attributes to the current buyer. Call this the "relevant set".
17 Given this, the probability $\text{Hit}(x, \text{Ad})$ that a user with profile x will hit an advert with profile
18 Ad can be computed as the ratio $\text{Hit}(x) = \text{Num_Hit} / \text{Num}$, where Num_Hit is the number
19 of users in the "relevant set" that responded to the advert, and Num is the total number of
20 users in the relevant set.

21 Although subject to a certain level of necessary uncertainty, because buyer behavior cannot
22 be predicted perfectly, ultimately it is useful to predict with as high an accuracy as possible
23 whether a buyer will respond. Given a probability $\text{Hit}(x, \text{Ad})$ that an individual buyer will
24 respond to an advert, a vendor can define a bidding policy. The policy maps $\text{Hit}(x, \text{Ad})$ to a
25 value to bid for the right to target a buyer.

26 Assume in this section that the profiling system places buyers into one of several "classes"
27 of buyers, $C1$ to Cn , given an advert with profile Ad , where each class has an associated hit
28 rate, i.e. $x \in C1 \Rightarrow \text{Hit}(x, \text{Ad}) = \text{Hit}(C1, \text{Ad})$. Furthermore, assume that the system also
29 predicts the average amount spent by a user that hits on the advert, or the average value to
30 the vendor from a hit, this can again depend on the class of buyer, and can be determined
31 within a profiling system based on historical information about this advert, or about similar

1 adverts. Let **Rev(C1, Ad)** denote the value to a vendor with advert **Ad** of getting a hit with
 2 a buyer in class **C1**. Now, the vendor can determine its expected value for placing an advert
 3 to a user in each class, for example using the computation **Val(C1, Ad) = Hit(C1, Ad) *
 4 Rev(C1, Ad)**.

5 The system of vendor-side advert competition works as follows. Once a user hits a web page
 6 with an iaworthit-targeted advert, a choice of adverts are pushed to the user's client from
 7 the ad server. Each advert is represented as a three-tuple:

8 (link to location of advert, bidding function **Bid(hit, rev)** , profiling function
 9 **Profile(x, Ad)**, value function **Rev(x, Ad)**)
 10

11 The profiling function **Hit(x) = Profile(x, Ad)** is used to place a buyer into the appropriate
 12 class of buyers, and compute the hit probability for a buyer, **Hit(x) = Hit(x, Ad) = Hit(C,
 13 Ad)**, where **x** is the profile for the buyer, **C** is the class that the buyer's profile places it
 14 within, and **Hit(C, Ad)** is the probability that a buyer in the class will respond to the advert.
 15 A possible implementation of the profiling function is to encode it using prototypical cluster
 16 centers for a buyer in each of the set of buyer class types, and then assign the buyer with
 17 profile **x** to the cluster that is closest (in some well-defined metric) to the cluster type. The
 18 metric does not need to be linear in each of the dimensions of a user profile, and in particular
 19 would be expected to be place more weight on terms that are important to the hit rate in a
 20 particular class of buyers.

21 As described, this method has the following useful characteristics:

- 22 a) the hit probability for a buyer is computed at the client machine, using the profile
 23 that is stored at the client for the buyer. The client machine does not need to
 24 release the user's profile, and the **user's privacy is assured**.
- 25 b) The information necessary to compute the hit probability can be encoded in space
 26 **linear in the number of clusters**, which is much more efficient than passing
 27 information about every relevant user profile to the client machine.

28
 29 Without computational/communication restrictions one might pass historical information in
 30 the form of (Profile, Hit/Miss) pairs for users that have been shown the advert, or shown a
 31 similar advert. A nearest-neighbor algorithm could then be used to find the relevant set of

1 profiles for a new profile **x** and compute the expected **hit-rate** from the ratio of users with
2 similar profiles that historically hit/missed the advert.
3 The cluster centers approximate this solution, representing the average hit-rate of “close”
4 profiles so that the client machine can simply locate the closest cluster center and use that as
5 a proxy for the probability that its user will hit the advert.
6 A similar technique can be used to compute the expected revenue from a buyer with profile
7 **x** that hits an advert. In fact, this information can be computed using the same method,
8 simply by associating an expected revenue with each cluster center.

9 **3.2.2 Client-side Advertising Reverse Auction**

10
11 The client implements an auction for the right to show an advert to a user. The auction is a
12 Vickrey auction between all the adverts that are passed by the ad server to the client. The
13 Vickrey auction (Vickrey, 1961) is a second-price sealed-bid auction. Given bids **b1, b2, ...**
14 **, bn** the auction sells the right to show an advert to a buyer to the vendor that bids the
15 highest value for the value of the second-highest bid. The Vickrey auction is useful because
16 it is truth-revealing. The optimal strategy of each vendor is to bid its true value for showing
17 an advert to a user. For example, with a profit-margin of 20%, it is optimal for a vendor to
18 submit **bid = 0.2 * hit * rev**. If accepted, it will pay at most **bid**, and it will pay only
19 enough to shave the bid of its nearest competitor. The auction is sealed, so competitors do
20 not ever see the information in failed bids. Cryptographic techniques can also be applied to
21 ensure that the bids are not inflated by the auctioneer.

22 A user can also define an acceptance function in the auction, which represents her
23 reservation price to view a particular advert. Assume that the reservation price depends on
24 the hit rate **hit** of the advert, and represent the price as **Accept(hit)**.

25 The client-side advert auction runs as follows:

- 26 1. For every advert **Ad**, Compute the hit rate of advert **Ad**, given the profiling
27 function **Profile(x, Ad)** and the buyer's profile **x**, and the expected revenue from
28 showing the advert to the user, **rev**.
- 29 2. For every advert **Ad**, with hit rate **hit = Hit(x, Ad)**, compute the acceptance level
30 **accept = Accept(Hit)** for the buyer and the bid for the vendor with the advert, **bid**

- 1 = Bid(**hit**, **rev**). If **accept** > **bid** then reject the bid for advert. If no adverts
2 remain, jump to step 7 (in this case no advert is shown to the user).
- 3 3. Given the bids that are not rejected by the client, sort them in order of increasing
4 value, and insert a “buy” bid from the buyer equal to the value of its acceptance
5 level for the advert with the bid of highest value.
- 6 4. Accept the advert with the greatest bid, and charge the vendor the price of the
7 second-highest bid.
- 8 5. Fetch the graphics and URL for the advert, and display the advert to the user.
- 9 6. The client monitors the actions of the buyer, and records (to be later transmitted to
10 the iamworthit server) whether the buyer responds to the advert.

11

12 The buy bid in Step 3 will be less than the bid for that advert, because the advert was not
13 rejected in step 2. However, this buy bid might be greater than the value of the second-
14 highest bid for an advert, and is required to make sure that the price paid by the vendor that
15 wins the auction is greater than the buyer’s acceptance level. We make sure that the user
16 cannot cheat by bidding just below the highest bid received by requiring that the user states
17 his/her reservation value before the value of bids are revealed.

18 A more general system for dynamic customized advertising might allow a vendor to specify
19 a bid for each type of buyer, and also a maximum budget, so that the vendor that places
20 adverts can maintain control over its spending.

21 We might also allow users to specify in their advert acceptance policy how many adverts
22 they are prepared to receive a day, so that they are not inundated with too many adverts,
23 even if they receive financial compensation.

24 **3.2.3 Numerical Example**

25 The advert auction message from the ad server arrives with a choice of three adverts, **Ad1**,
26 **Ad2**, and **Ad3**. Each advert is associated with a profiling function, a bidding function, and a
27 value function. The client machine computes the hit rate for each advert, based on the
28 profiling functions and its local profile for the buyer. Suppose **hit1** = 30%, **hit2** = 5% and
29 **hit3** = 15%. The client machine also computes the expected revenue if the buyer hits an
30 advert, **rev1** = \$2, **rev2** = \$8, **rev3** = \$3. The hit rate and revenue are used to compute bids

1 for each advert, using the bidding function. Suppose that the bids are **bid1** = \$0.50, **bid2** =
2 \$0.70 and **bid3** = \$0.40.
3 Now, the client also computes the acceptance level for each advert, based on the hit rates
4 predicted within the iamworthit system. Suppose **accept1** = \$0.30, **accept2** = \$1.00 and
5 **accept3** = \$0.30. The bid for advert 2 is rejected because it is below the accept value. The
6 bids for adverts 1 and 3 are accepted. Now, the auction is constructed with bids **bid1** and
7 **bid3**, and **accept1** because that is the accept-value for the bid with the highest value that is
8 not rejected. The auction takes the bids (0.50, 0.40, 0.30), and sells the right to advertise to
9 the user to vendor 1 for \$0.40 (the value of the bid from vendor 3).
10 Finally, the client fetches the graphics and URL link information for the advert from vendor
11 1, and transfers payment from the vendor to the user for the right to show the advert.

12 **3.2.4 Tuning an Advertising Strategy**

13
14 Although the optimal strategy for a bidding agent in a single Vickrey auction is to reveal its
15 true value for the right to show an advert to the user, the vendors are bidding over a number
16 of different auctions, and might have a limited advertising budget.

17 *Selecting appropriate users.*

18 One good approach is to start with a reasonable policy and then adjust it dynamically, based
19 on feedback received from adverts using techniques from reinforcement learning (see [RN
20 97] for an introduction). Off line simulation with methods such as Monte Carlo simulations
21 can provide good initial policies, based on information about a simulated population of
22 users, which could be provided anonymously by the system of Secure Data Interchange.
23 Users can be provided with incentives to reveal information anonymously about their
24 acceptance functions, to allow this type of off line modeling.

25 A vendor can tune an advertising policy with a random sample of users, simulating the
26 auction that runs on a user's client machine. Metrics such as the average number of times
27 that an advert is shown to a user of each type, and the average amount that the vendor pays
28 to show the advert provide information to allow optimization. For example, the analysis
29 might show that although a particular class of buyers are most likely to hit an advert, there is
30 also a lot of competition to show adverts to buyers in that class, and the average cost to show
31 an advert is high. In this case a vendor can conclude that it is more cost-effective to

1 advertise to buyers in cheaper but less relevant classes. The analysis can also be used to
2 check that the average buyer is not setting an acceptance level higher than the vendor's bid,
3 which again would indicate that the policy should target a different set of buyers.

4 *Selecting appropriate adverts.*

5 The system that we have described can be extended to allow a vendor to select an advert to
6 display to a user based on the type of profile of a user. For example, a vendor can pay for the
7 right to be one of N vendors that compete in an auction for the right to advertise to users that
8 hit the web page of a particular vendor. We can allow a vendor to not only tune its bid to the
9 profile of a particular user that hits a web page, so that only users with a good fit with the
10 service offered receive the web page, but we can also allow a vendor to tune the advert that
11 it shows to a user.

12 We allow a vendor to submit a number of different adverts and bid functions via the
13 advertising network server. The client-side auction can be expanded to allow multiple bids
14 from each vendor, where each vendor can submit a number of bids, but only pays the second
15 highest price bid from another vendor. Again, the auction is truth-revealing for a vendor.

16 **3.2.5 Automatic Advert-Replacement Systems**

17
18 Although there are potential copyright violations in any system which alters the content of
19 information published by a vendor before displaying that information to a user, we comment
20 that the above described system can be applied even with vendors web pages that do not
21 subscribe to the advertising network in SDI. Adverts can either be replaced, or added in
22 separate windows. This enables any user that subscribes to SDI to receive personalized
23 adverts and revenue from advertisers, irrespective of whether or not pages are within the
24 SDI network.

25 The key problem in ad replacement is to identify an advertisement within a page, i.e.
26 information that is not related to the core purpose of the page. At present adverts may take
27 one of two forms:

- 28 (a) They can be statically, or dynamically, generated at the web server of the web
29 page that the user hits, and pushed to the user directly at the same time as the
30 general content.

1 (b) They can be pushed to the user in parallel with the other non-advertising data on a
2 web page, for example with a link to an advertising network.

3
4 The adverts in (b) can be identified by tracking ad server URLs that are embedded in the
5 source code of web pages. Case (a) is more difficult because there is no identifying URL to
6 indicate the content of the section of a web page. However, there are solutions to advert
7 identification in these cases, for example via regular expression identification and the
8 location of ad banners, as described in www.junkbusters.com and the “block files” that
9 allow a proxy server at junk busters to strip adverts from pages. Another technique is to
10 identity URL’s that allow a user to click to another domain from within the current page.
11 These might likely be adverts, consider for example, a link to www.buyanewcar.com on the
12 New York Times web page.

13 Another variation applies to television commercials and other broadcast medium, especially
14 as the data storage capacity of local set top boxes increases. For example, a code is
15 transmitted by national networks to cue local adverts from local network stations, allowing
16 the replacement by personalized adverts from advertisers in the SDI ad network. In the
17 future it is also likely that we will use digital storage devices to receive entertainment and
18 television programs that are broadcast, storing the information for future viewing, or even
19 slightly delayed “on demand” viewing. With digital media and storage, it is possible to
20 replace adverts inserted into programs by the producers with SDI targeted adverts, either
21 with or without the cooperation of the source of the information.

22 3. Additional Applications

23
24 1. Use of Real Time and Anticipated Vehicle Location Data to Provide Real Time
25 Traffic Reporting and Predicted Traffic/Congestion Modeling

26
27 a. Real Time Traffic Reporting

28
29 A simple but useful application of LEIA involves on an opt-in basis real time
30 transmission of a sample of driver’s location information to a regional traffic-reporting
31 bureau. Though existing prior art approaches are reasonably efficient, there are common

1 circumstances in which congestion may have just occurred (e.g., following an accident)
2 or wherein congestion has recently cleared up and such up to the minute information is
3 not available to the traffic bureau. Moreover, this approach could be extremely
4 advantageous in providing optimal navigational assistance for drivers wishing to
5 optimally reduce driving time to a particular destination, which she/he submits to the
6 system. The present system would consider the comparative length of different routes
7 from the user's present location to the destination in combination with the anticipated
8 average speed based on that of other vehicles traveling those particular routes. This
9 enables real time recommendations to be presented to the user, based on this real time
10 data.

11
12 b. Predictive Modeling of Traffic Patterns Based upon Real Time Location Data from
13 Vehicles

14
15 It is possible to apply certain statistical techniques in order to predictively anticipate
16 traffic patterns for some marginal period of time into the future based upon real time
17 location data regarding the present case sample of vehicles. One may consider recent
18 past traffic patterns compared to present traffic patterns (i.e., change in speed as a
19 function of time by segment of highway as well as absolute and rate of change in traffic
20 volume). There are pre-existing models which are able to anticipate the corresponding
21 effects of congestion at the moment in which it begins to occur which utilize these
22 variables. This model must also consider the rate of change in average speed over each
23 segment as well as the predicted impact of changes in traffic volume from feeder
24 highways based upon changes in volume occurring dynamically thereupon. It is a key
25 objective of the navigational assistance feature to utilize future predictive traffic models,
26 in order to optimize the reduction in transit time for the most number of users possible,
27 i.e., considering the volume of traffic which is predicted on a segment by segment basis
28 in combination with the traffic models for the impact of these volumes upon transit time
29 (including the rate of increase thereof). With this data, the system can determine, in
30 advance, on an individual driver basis, which route for a given requested destination, will
31 provide the shortest drive time for the user, while insuring that a near optimal transit

1 time is achieved, across all segments of the highway system, as a result of each individual
2 system recommendation to each driver. For example, an anticipated compounding
3 congestion problem would warrant the vast majority of vehicles to take alternate routes
4 until the congestion or predicted congestion is alleviated in that segment.

5
6 2. Use of LEIA for Creating Traffic Models for Purposes of Mapping Commercial
7 Industrial and Residential Real Estate Market Opportunities –

8
9 A very useful application of LEIA involves the collection of traffic pattern data on a
10 time-specific basis (the day of the week and time of day) for the traffic, as it passes each
11 piece of real estate. An electronic map which is ideally Web based (potentially
12 nationwide is generated and constantly updated based upon this data. Additional
13 information may be provided which may include (but is not limited to):

- 14
15 a. Origin and destination information of the vehicular traffic (as captured by LEIA)
16 which may, especially if correlated with time, suggest the nature and context of the
17 driver's activities, e.g., rush hour traffic, errand traffic, etc. It may be useful to factor in
18 the type of neighborhood the vehicle returns to every night, the type of commercial or
19 business entity she/he drives to work to each day, etc.
- 20
21 b. Other activity-related clues which the user is willing to release, e.g., devices interacted
22 with, content interacted with or transmitted information, etc., which may provide insights
23 into the mind-set of which users tend to experience when in the vicinity of the real estate
24 property.
- 25
26 c. User Profile Data – Aggregate purchase and content affinities as well as price
27 elasticity data (gleaned from purchase statistics) could be very useful information to
28 commercial real estate developers and purveyors. Users with the right profile and a
29 receptive mindset are of particular interest.

1 It should be noted that the present system may be extended to residential real estate. E.g.,
2 what types of jobs (such as quality of jobs) do local commuters have? What are their
3 numbers? How far do they commute (particularly if they tend to commute further than
4 the present real estate site)? Do their commuting routes tend to pass the current potential
5 real estate site? The last three questions would also be particularly relevant as well for a
6 prospective industrial real estate development opportunity.

7
8 Industrial real estate developers also may be interested in mapped models of real estate
9 depicting the professional and known likely educational characteristics of the associated
10 local residents in that region? What are the other businesses at which they work? (If
11 available) what are their particular positions/responsibilities?

12
13 3. Use of LEIA for Creating Traffic and User Profile Models of Traffic Passing
14 Billboard Sites and Providing a Map of Such Information on an Available Billboard Site
15 Basis –

16
17 The presently described techniques for providing dynamically updated informational
18 maps containing detailed statistical data regarding vehicular traffic passing real estate
19 sites can be further extended to similar maps of interest to advertisers which contain
20 locations for available highway billboards. The present system further provides Web-
21 based access, which enables advertisers to make reservations and purchases of such
22 billboards. In one preferred variation, an economic model is deployed to optimally price
23 the billboards. I.e., a varied representative sample of each type of billboard sharing
24 similar traffic/user profile features with others is auctioned for this purpose. In another
25 variation this on-line auction model is deployed for all billboards available by the system
26 in order to provide a novel service to advertisers which is a “billboard auction” site.

27
28 4. Use of LEIA for Enabling Drivers to Identify and View Sales Opportunities and
29 Offers Associated with Physical Objects in the Surrounding Environment Utilizing
30 Heads-Up Display Technology

31

1 The present system is an extension of the virtual tag methodology in which physical
2 objects contain meta-data and in which the location of such objects with respect to the
3 user is determinable either by transmission of a signal, which contains the exact physical
4 coordinates of the object or a scheme by which this information may be pre-loaded,
5 which is typically from a server, which contains such information (thus “non-wired”
6 objects may be identified accordingly) to the user’s device, e.g., as she/he physically
7 travels in the vicinity of such items. In either case, the present physical location of the
8 object must be known in addition to that of the user (in the case of fixed objects, of
9 course, the dynamic location tracking techniques are unnecessary). A primary
10 commercial application of the present scheme applies to purchasable items. The user
11 may submit his/her interest profile to SDI for purposes of being dynamically notified and
12 disclosed of purchasables which match his/her user profile or specifically requested
13 search criteria. The metadata may include a brief description and any additional level of
14 information which the seller may be willing to disclose to that particular user. The seller
15 may possibly request details regarding price elasticity of the buyer (purchase appetite)
16 before quitting any discounts to the existing price. Prime example applications of the
17 present scheme may include:

- 18
- 19 1...Revealing used car sales opportunities to drivers (by actually flagging relevant
20 vehicles which the driver passes (or which pass the driver);
21
- 22 2. Real estate and home sales opportunities (as well as apartment and commercial real
23 estate rental/leasing opportunities;
24
- 25 3. User-user introductions in which one or both users may be extremely interested in the
26 other based upon matching criteria within their associated social, information and/or
27 professional desires and assets.
28

29 The preferred underlying technology for the present heads-up display system involves a
30 technique for tracking the orientation of the user’s head as well as the direction of his/her
31 pupils (it may be possible, however, to implement the system by detecting pupillary

0969098-102700

1 direction based solely upon the position of the pupils relative to one another, using a 3-D
2 optical tracking device. The iris of the user may also be utilized as a bio-metric identity
3 of the particular user (see issued patents assigned to IrisScan Corp.) In the preferred
4 implementation, the head-up projection device projects the virtual flagger upon the
5 portion of the windshield or window(s) which is in direct line of site between the user's
6 pupils and the physical object of interest. The technique may utilize heads-up windshield
7 display technology and/or remote projection (e.g., for projection upon the side windows).

8
9 In another more advanced version of the system, the holographic projection system using
10 micro-mechanical techniques may project the information directly upon the user's retina.
11 This technique is detailed in co-pending patent application entitled "Remote Retinal
12 Imaging Projection System" The variety of other potential applications of the present
13 technologies are also conceivable which include:

14
15 **SYSTEM FOR PROVIDING REMOTE ACCESS AND MODIFICATION**
16 **CAPABILITY OF DOCUMENTS AND E-MAIL BY A SENDER**

17
18 The present system involves the use of a technique which enables the sender of a
19 document, upon permission of the recipient, to provide subsequent updates and revisions
20 to that document remotely and automatically in the absence of the recipient. Utilizing the
21 computer's modem and e-mail delivery system, these updates could be provided remotely
22 by the original sender (who alone can gain appropriate access privileges through a unique
23 access code or authentication). The sender may also be able to identify whether or not
24 the old version had been accessed by the recipient and possibly even at the level of the
25 particular segment, or segments in which the specific changes had been made. (This
26 capability could also be two way in which certification of not only receipt, but also access
27 particular portions of the information by the intended recipient could be positively
28 confirmed by the sender however, subject to approval by the recipient). If the recipient is
29 amenable, such receipt confirmation could also be automatically time stamped by a
30 secure trusted agent on the recipient's machine. The system could also perform the
31 appropriate modifications in different drives on which the document is stored. If the

1 appropriate modification was performed on the hard drive, but not a copy on the
2 floppy(s), the recipient could be notified and prompted to insert the floppy(s) for
3 appropriate updating until all versions of the original document are updated. It should be
4 noted that such latter functionality would even be of independent utility to standard
5 modifications as performed with word processing systems. The above same features
6 could also be provided for e-mail as well(in a variation, time stamped receipt could also
7 be applied to voice mail as well). The recipient if he/she had not approved the remote
8 modification in advance, the sender could request the access/modification privilege of the
9 recipient via e-mail or instant messaging service for example. The recipient may also be
10 able to identify a digital time stamp for the relevant portion(s) of the document (or e-
11 mail) as modified/updated by the sender. In a variation, XML metadata containing
12 digital time stamps could also be utilized for the standard modifications, which occur
13 throughout the document, i.e., the nature of the modification, i.e., the addition, change or
14 deletion, the time it was performed, on which machine, under which password and if
15 relevant by which remote user. Outside of the context of the present remote modification
16 system this function could be integrated into a standard word processing system with
17 standard comparison check function. Such digital time stamps would help in proving
18 identity and date of authorship at a detailed level.

19
20 In a variation of the present system, a user could also receive automatic notification
21 if/when a recipient has accessed (or provided an associated modification to) a document,
22 portion thereof, e-mail or voice mail. This would require the recipient to provide
23 approval of the above privileges for the sender in response to a request prompt. In
24 another variation of the present system, the provider of a URL may receive time-stamped
25 verification of a user having accessed a URL (which may request a uniquely identifying
26 access code or other authentication or the Web page may be accessed by that user who is
27 provided with a completely unique URL, thus verifying access by that user for remote
28 access, modification and associated time stamping. If the URL is not under control of the
29 requestor of the verification the accessor's client or Web-centric script (associated with
30 his/her access account) could notify him accordingly with similar time-stamped
31 verification of access. In another related application of the present capability, the

1 5. Verify and, if desired, notify requestor (e.g., employer or teacher/parent) of accessor
2 (employee or student/child respectively) of accessing, generating or conversing
3 (acoustically/telephonically or textually) certain content which falls outside of a
4 certain defined set of content clusters which are deemed "appropriate" within certain
5 time periods (e.g., work-time such as when the user is "clocked-in", homework time,
6 etc.) or in general. And/or verify or notify accessor of certain clusters, which are
7 accessed or generated within certain content clusters. The above cluster based
8 identification rules may also be definable by not only time of day or day of week, but
9 domain, e.g., e-mails and phone conversations of a "social" context or style are
10 inappropriate while pornography is inappropriate during work time for employees or
11 anytime for students or children. The present "conversation" based variation, may
12 even be implemented in conjunction with a passively collected acoustic monitoring
13 via the client's microphone of an employee's, student's or child's spoken
14 conversations in/around his/her (or a) client machine. Identity may be determined in
15 conjunction with associated voice print information and bias results to documents,
16 which match that style in accordance with the appropriate weighting criterion
17 matching the user profile.

- 18
19 b. Give the user a set of selection criteria to bias the search (or recommended
20 navigational links) by those styles, which s/he desires presently.

21
22 In addition, the present system could further actually use style as an additional means for
23 identifying documents, which may be relevant to a search, e.g., determine what style(s) a
24 preponderance of document in the present search tend to possess, then add this criteria as
25 an additional weighting criteria to determining document relevance (or finding new
26 potentially relevant documents). I.e., as part of the information retrieval (attribute
27 cluster) model, use "style" as an additional attribute for and similarity to other documents
28 and queries.

29
30 System for Mining User Reactions and Responses to On-line Media by User Profile
31 Characteristics:

1

2 As described, are some primary examples of how the present technique can usefully be
3 deployed. User profiling may also be usefully applied for purposes of collecting feedback
4 about various types of user reactions and responses to various on-line stimuli (e.g. general
5 Web and e-commerce sites to various products such as movies, music, interactive content,
6 advertising news, interactions with other users etc. A somewhat related methodology was
7 disclosed in the parent patent involving a technique for allowing users to rate vendors
8 according to a variety of criteria and enabling the vendors (e.g. for marketing purposes)
9 and/or users to observe statistical correlations between the user profile characteristics and
10 ratings (using data mining techniques). In a simple example, application of the present
11 methodology, the users viewing a Web page (containing any of the above suggested
12 informational types) are allowed to provide their personal comments about the page or about
13 their own experience in viewing the page. Informational retrieval and statistical NLP
14 techniques may be used to cluster (using standard clustering techniques) both the comments
15 (by similarity of their content profiles) as well as the users (by similarity of their user
16 profiles). In the first instance, the comment cluster exemplars are extracted, used as the
17 comments which most closely exemplify each of their associated clusters and thus presented
18 to vendors and/or users. The aggregate (average) of the user profiles associated with that
19 cluster, the user profile of the user who provided the exemplar comment or the subset of
20 user profiles which characterize subclusters of the comment clusters (wherein the
21 subclustering routine is based upon the profiles of users within each comment cluster) are
22 also presented to the users or associated vendors in conjunction with the exemplary
23 comment most closely associated with that stereotypical profile of users. In a variation,
24 which could be an automatically selected alternative if the exemplar fails to provide a clear
25 representation of the various comments in the cluster (if further subclustering does not
26 adequately achieve this objective), it may be possible to "combine" these various comments
27 which are closest to the clusters' centroid by identifying those portions of the comments
28 which are similar in meaning and those portions which are different. For similar portions,
29 by preferentially utilizing the counterparts which are closest to the centroid and adding to it
30 those portions which are different in piecemeal fashion, the desired objective may be
31 achieved.

1

2 In a variation, the different types of comments may be clustered and accessible to users via a
3 hierarchical cluster tree used to create a menu of automatically labelled clusters (see parent
4 patent detailing this general method). Users may access such menus by combining querying
5 (of user or content attribute with menu navigation).

6

7 It is of course reasonable to combine the technique in the parent patent application for
8 statistically correlating user profile attributes with ratings associated with the users
9 possessing those attributes.

10

11 It is also reasonable to apply the present technique to video (or streaming video content)
12 whereby during the course of viewing the content relevant survey questions are presented to
13 users and if desired, the correlation statistics of the results with the associated user attributes
14 are presented to the content owner and/or (preferably) also future viewers. In a novel and
15 advanced variation, time shifted video technology would enable the ability to capture audio
16 or full video/audio versions of the comments of the user.

17

18 In a very novel application, it is possible upon permission of the associated users, to
19 automatically construct menu trees (in accordance with the above method as disclosed in the
20 parent case) of live voice or text chat in real-time dynamic fashion. Spoken conversations
21 may be automatically profiled using acoustic speech-to-text methods. In accordance with
22 the methods for automatic construction of virtual communities (in this case for dynamic
23 chat) as disclosed in the parent case, these communities can be constructed automatically
24 and on-the-fly based upon similarities of chat content, queries, navigational (content)
25 selection from automatically constructed menu and/or user profile similarity or attributes of
26 the profiles of the users which may be selected. Thus, in the former case a user could in the
27 case of a pull down menu gain access by title of key phrase (which are ideally automatically
28 and dynamically created and updated in accordance with the content of dynamic spoken or
29 recorded conversation spoken or typed in real time and which may be further dynamically
30 categorized and re-categorized within a hierarchical menu structure format. Although, the

1 word(s)/phrase(s)(again, so long as the user's profile is within the disclosure policy of the
2 transmitting user. If a willing recipient so desires for both the chat and instant messaging
3 variations, s/he may receive a copy of the text or voice transmission or if the user is not able
4 to access it in real-time or otherwise. Conversely, the user may as a sender wish to send the
5 transmission to

6 1. All users who have indicated potential interest in the contents (and/or profile or identity
7 of the sender) or users who match a particular user profile (e.g., interest or proficiency such
8 as which is of measured similarity to the contents of the transmission) and/or
9

10 3. Users who are in the process of engaging in a user-user (or multi-user) dialog, which is of
11 measured similarity to that of the transmission.
12

13 An example application of the present system includes, for TV viewers, the ability to gain
14 access to different types of feedback from other users who are also presently viewing or had
15 previously viewed a particular TV program, e. g., a comedy, a news story or political speech
16 in which the present technique could be used effectively as a filter allowing certain types of
17 user or user attributes to reach the user and others to be suppressed or squelched. In the
18 latter example, a user who considers herself to be a female liberal may provide settings
19 during a campaign speech by a conservative Republican speaking out against abortion to
20 listen to acoustic/verbal gestures by other liberal females advocating abortion). Audible
21 comments may be unfiltered and heard by the user, if desired.

22 If the program is a re-broadcast, more elaborate filtering capabilities are conceivable
23 involving statistical analysis of the spoken language content (such could be conceivably
24 performed for real time live information, however, at the expense of a slight delay). As
25 suggested, it may be also possible to observe textually (e.g., through speech to text
26 techniques) or hear spoken conversations as they exist between individuals possessing
27 desired attributes, containing content characterized by key words or phrases from a menu
28 selection or provided by/between a desired individual or individuals respectively.
29

30 In another application, users viewing video or streaming video content may upon their
31 permission, agree to disclose their user profiles and be acoustically monitored such that

1 traits are also used. In the present extension, we first suggest a primary system
2 methodology in which extremely detailed information is collected and aggregated into
3 the user profile particularly pertaining to the user" psychological profile (user
4 preference/interest information demographics, etc. are also useful data in that there are
5 likely correlations which exist with certain psychological attributes of the user). In the
6 clinical field of psychoanalysis, much information regarding the psychological and
7 psycho-pathological characteristics of the user is determined, however in compliance
8 with the confidentiality requirements of the doctor-patient privilege, most if not all of this
9 information remains isolated and never accessible by a large-scale statistical database for
10 purposes of cross-correlation of psychological characteristics, phenomena, psycho-
11 pathology (as well as other clinical pathology), as well as behavior, interests, preferences
12 and more specifically identifiable behavior patterns as well as notable life experiences.
13 Part of the answer to this dilemma may well lie in the application of secure data
14 disclosure to SDI, the pseudonym proxy server and (for extra security) randomized
15 aggregates for purposes of harvesting the statistical value within the data and (if needed),
16 interacting with the user while maintaining completely secure and private individual user
17 data. The present system suggests that there may be much greater accuracy which could
18 be achieved in the profiling of individuals for use in a number of applications such as the
19 psychological modeling of patients for use in the diagnosis, treatment, prediction
20 (including predictive treatment) of psycho-pathology, the matching of "similar" patients
21 together (for group therapy and/or pseudonymous electronic or physical mail
22 correspondences) for providing support, inter-patient communication, counseling, etc.
23 (for which present traumatic life experiences, if relevant, may be quite applicable as
24 well), as well as the extended/improved modeling/prediction of user interests/preferences
25 enabled by SDI.

26 Another correlated data type could be employee file databases through which correlations
27 and predictions of employee behavior and performance, recommendation as to remedies
28 for existing problems and performance enhancing tactics again could be achieved.
29 Unfortunately not all data is likely to be accessible for all users (e.g.,
30 browsing/transaction histories are likely to be available but often not information
31 regarding a user's life experiences or psychological history). To fill in this missing data,

1 rapid profiling techniques could be applied (see parent patent). Key aspects of a user's
2 psychological profile for example, could be acquired, anything from their personal
3 experience, beliefs, fantasies, habits, personality traits social behavior, etc., for
4 preventative life style patterns. In an obvious example application, an individual could
5 be very accurately profiled, psychologically in order to predictively determine likely
6 strengths and weaknesses in his/her psychological and psycho-social aspects and thus
7 determine recommendations for preventative life style patterns (which could be provided
8 by professionals and/or determined statistically or confirmed/disproven in order to
9 enhance potential strength and avoid or remedy potential or existing weaknesses (in as
10 much as the pseudonymous statistical database would, of course, also contain the results
11 of certain lifestyle therapeutic or pharmacological treatment approaches). It is believed
12 that in the future, genetic sequence data which could be stored in the pseudonymous user
13 database could also provide valuable clues as to the complete medical, psychological and
14 psycho-social make-up of individuals. Another potential methodology could be to
15 present particular emotion-evoking stimuli as a textual, verbal (or most ideally)
16 multimedia videographic or even emersive virtual reality) presentation, capture details
17 about the user's emotional response, comments, gestures, e.g., crying, laughing repulsion,
18 mortification feeding of physical illness, etc., and perhaps even physiological feedback, if
19 accessible.. With such large quantities of such sensitive data regarding individuals, SDI
20 offers users a very powerful and compelling solution by which user's SDI agent can for
21 any given application or circumstance select only those potions of his/her profile which
22 will ultimately benefit the user. I.e., there are typically for most relevant applications,
23 e.g., employee hiring by corporate HR departments, health insurances,etc., extremely
24 crude metrics which are used in the determination and/or terms of a prospective contract
25 between such commercial entities and a user. There is, thus, considerable opportunity for
26 SDI agents of the user to analyze user profile data in light of pre-determined models of
27 what features and metrics are most significant in determining selection and/or terms of
28 service or a professional relationship with that user based on historical data (through
29 feedback) or publicly available informational models (e. g., as in the highly regulated
30 health insurance industry). In order to initialize the statistical model for the more
31 subjective example of employee selection and terms, the present system could , for

1 example, disclose to the employer "Mr..X has a psychological profile feature which is
2 highly correlated with a high aptitude for the present job which s/he is applying for and
3 past job experience Y is 50% more likely than the average employee for that position to
4 lead to a subsequent promotion and several "employee of the month awards"
5

6 Application to Identifying Similar Individuals Behaving Under Similar Circumstance 7

8 The present user profile information could also be extremely relevant for a novel Internet
9 based service by which individuals could pseudonymously release their profiles and/or
10 define specific situations and/or conditions (which could be actual, hypothetical or
11 combination thereof) and the system will identify other scenarios (of like pseudonymous
12 individuals and/or circumstances) to the general scenario and emphasizing and
13 prioritizing certain particular aspects thereof. Examples include identifying individuals
14 who had previously (or are presently) in the process of making a very similar professional
15 decision, interpersonal decision (such as regarding a present or a potential spouse).
16 Ideally, such interpersonal decisions would involve a very similar situation (such as a
17 decision), a very similar individual in the analogous position of the user (and ideally, if
18 possible other similar party(s) fulfilling the analogous role(s) as the present situation
19 which confronts the user. The relative importance (priority) of the degree of similarity of
20 various analogous roles in an analogous situation is determined in part by the type of
21 information which the user would like to determine. For example, determining what a
22 particular individual would do in a similar circumstance may depend upon how
23 dependent that user's action .or decision is upon the circumstances and conditions versus
24 another individual(s) (and which individual(s) and/or the two (or more) individuals
25 involved to which the context of the situation primarily relates. Of course, the nature of
26 the action in question may predicate the relative importance of one or more of the above
27 factors. These determinations are often quite subjective in nature. Requestees may also
28 (if the choice exists) be situations which are either of a historical nature (i.e., the situation
29 had occurred in the past) or of a present nature. In some cases in the latter case, is
30 preferable as the user may actively request, e.g., the party to which s/he directly relates or
31 is potentially affected by to provide advice and/or feedback or to what s/he would do

1 the user may be readily extended to predict or help to substantiate other statistical user
2 variables relating to the types of actions and content related profiles characterizing the
3 user's temporal behavior patterns. The following application of these techniques is used
4 to enhance the automatic meeting scheduler:

5
6 The identity of the task or request, the activity and content attributes of the task or request
7 profiles (target object profiles) may be developed utilizing the priority of that task
8 relative to the user. This priority is estimated from the priority of the task/request (and its
9 associated attributes) over other tasks/requests by the user. In order to better substantiate
10 the relative importance to the user as inferred regarding these activities, the system
11 should also display the scalar ratings which the user could adjust manually.

12 Nonetheless, there may be a degree of uncertainty in the user's intentions and LEIA's
13 inference of the user's intentions. Moreover, unanticipated changeability by the user
14 (uncertainty) may occur perhaps with increased frequency under certain variables like the
15 identity of the requestor, the time of day, context of the users other present activities or
16 when other meetings and obligations are pending which possess a relatively greater
17 priority with the user (i.e., if they often occur spontaneously). In order to take this
18 uncertainty factor into consideration, an overall statistical estimate may then be
19 calculated taking into account the predicted statistical probability of each party ultimately
20 being available for the meeting. This serves the purpose of both refining the automatic
21 scheduling process and informing would be visitors before hand as to the relative
22 importance that their prospective meeting is (or increasingly becomes as the meeting time
23 approaches) to the employee as well as the overall estimated degree of certainty of the
24 meeting. This feature may be particularly useful for impromptu visitors where
25 considerable uncertainty exists as to whether and to what degree the visitor is imposing
26 upon the employee's schedule.

27
28 The temporal context of user behavior (using temporal time series analysis in conjunction
29 with use of content analysis and user behavior assessment may be both important
30 predictors as to whether a prospective meeting of a disclosed purpose is worthwhile for
31 the employee based on present/future priorities and at which most likely times. By far,

1 its most valuable application , however is in its ability to perform content analysis and
2 user behavior assessment in real time in order to enhance LEIA's ability to determine
3 whether and to what degree a would-be unannounced visitor or telephone or intercom
4 callers unannounced contact is relevant to the employee's present activities and if not to
5 coordinate with the visitor's agent the time or times when it would be most relevant to the
6 employees anticipated activities for example there are certain activities during which an
7 employee doesn't wish to be usually disturbed, e.g. while reading the morning news,
8 taking a coffee break, calling home, responding to an important email message, ten
9 minutes before an important meeting. Often within an organization individuals will
10 attempt to speak with one another as such , the intended duration affects priority as does
11 the identity of the individual in the organization. Often this variable can be automatically
12 predicted by LEIA . In such a situation, the relative priority of the user's preexisting
13 priority tasks is weighted against the priority of the requested tasks prior to notification.
14 The above application can be readily extended as well to phone calls, in which the user
15 could be automatically prompted by LEIA to disclose his/her purpose of the call and/or
16 his/her identity if necessary. This application would be a useful extension to the
17 telephony variation of the email filter described in issued patent entitled "System and
18 Method for Electronic Identification of Desirable Objects in which the above additional
19 inputs about employees and visitors may provide additional useful feedback in
20 automatically determining the rules dictating how to best handle the call or forward it to a
21 more appropriate employee on behalf of the employee. If there is a certain degree of
22 uncertainty in the agent's priority determination or assessment of the task affected the
23 most appropriate rule the employee may observe the rule just prior to automatic
24 implementation. Or the action which it represents. The present value of the contextual
25 employee activity assessment method as disclosed, if extended to advertising could be
26 significant. For example the knowledge of the context of the user's present activities
27 could determine what type and when an advertisement targeted to an employee, e.g. a
28 decision maker will most likely elicit the best response (where the advertisement
29 matches the professional interest profile) is an ad which corresponds with particular
30 activities of the user (e.g., which relate in some way with the ad/promotion. As
31 discussed direct voice telephony and email dialogues may provide very useful clues

1. A system for exchanging data, comprising:

a first and a second party connected to the communication system, wherein each party has personal data, and each party has a disclosure policy to control dissemination of its data; and

2. The system of Claim 1, wherein the data exchanged by the secure intermediate party is excluded from being transmitted to a third party in accordance with the disclosure policies of the first and second parties.

ABSTRACT

1 A system for exchanging data includes a communication system, a first and a second
2 party connected to the communication system, wherein each party has personal data, and
3 each party has a disclosure policy to control dissemination of its data, and a secure
4 intermediate party connected to the communication system, wherein the secure intermediate
5 party exchanges data between the first and second parties in accordance with their respective
6 disclosure policies.
7

09699098-102200

Figure 1. Top-level SDI architecture.

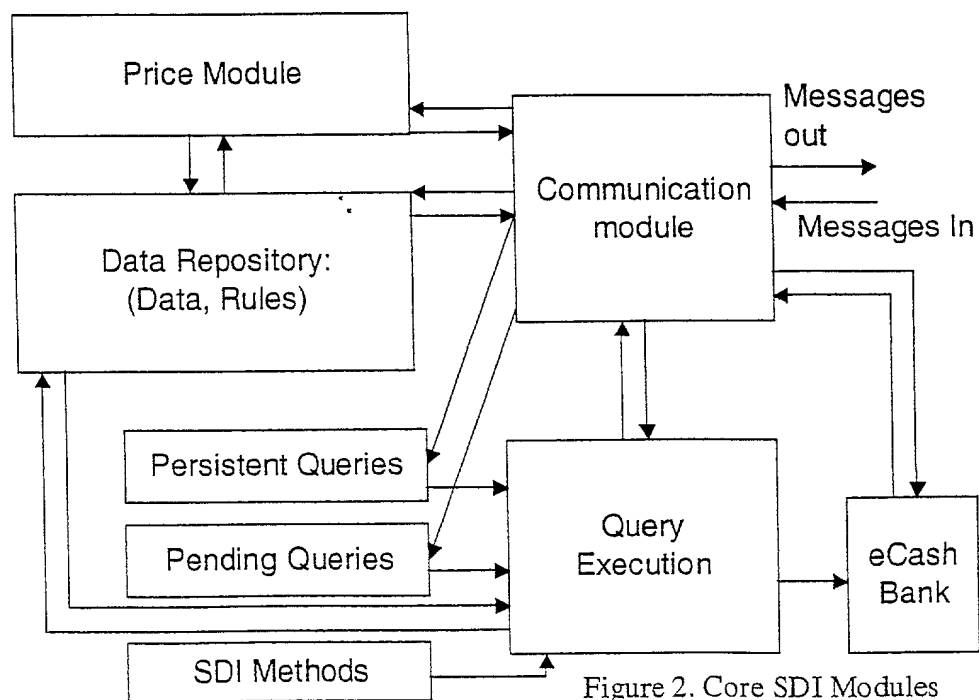
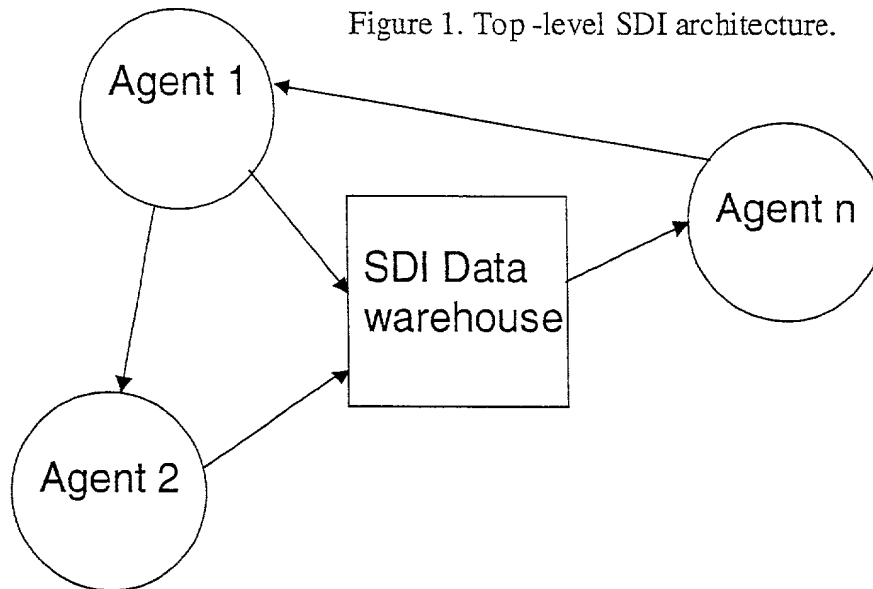


Figure 2. Core SDI Modules

002201 86066960

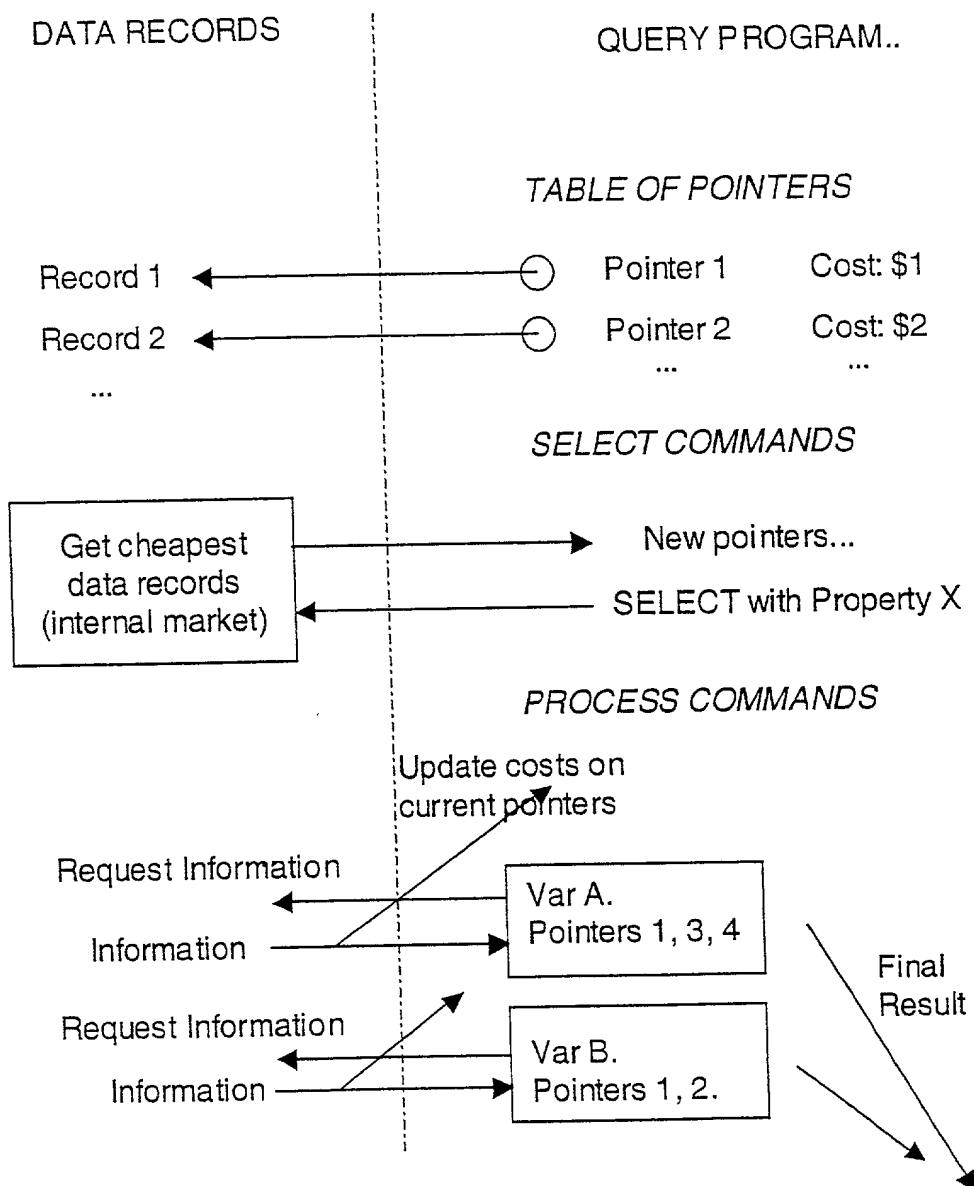
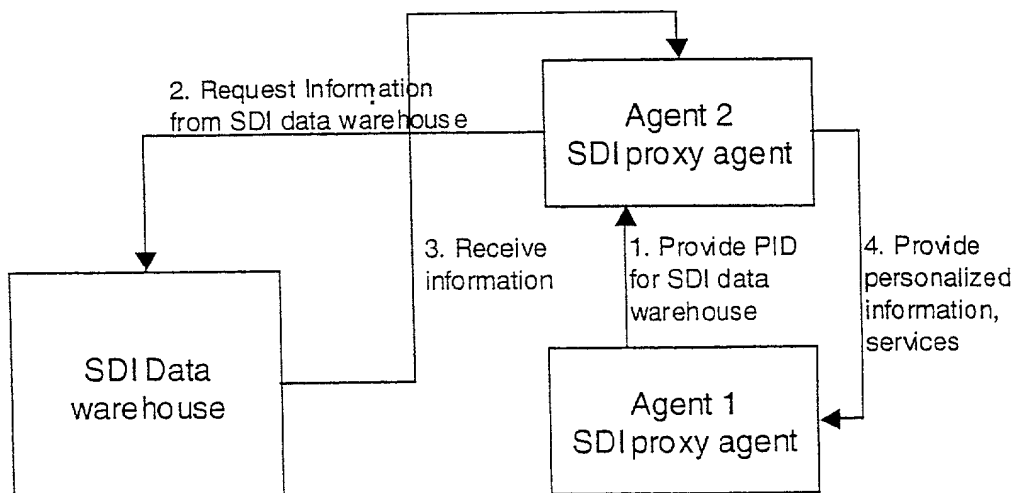
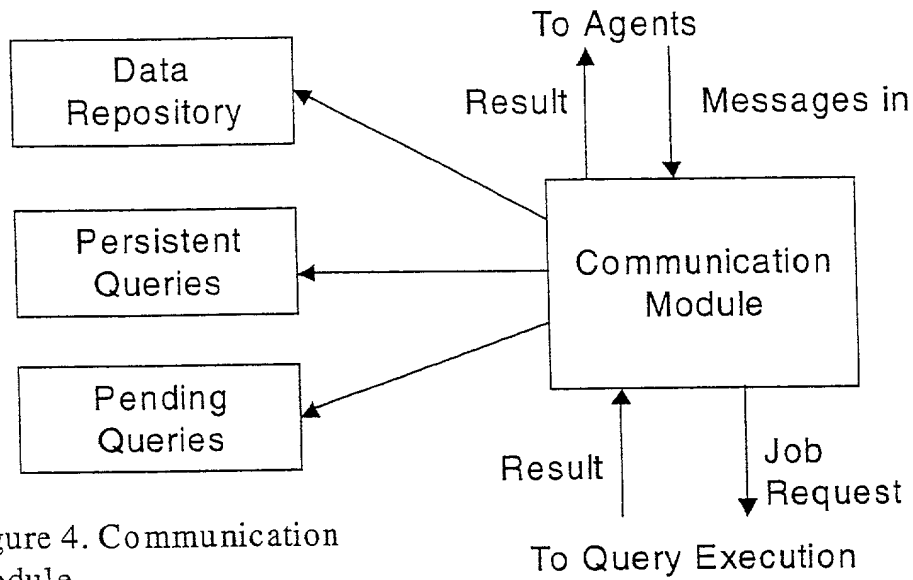


Figure 3: Query Execution



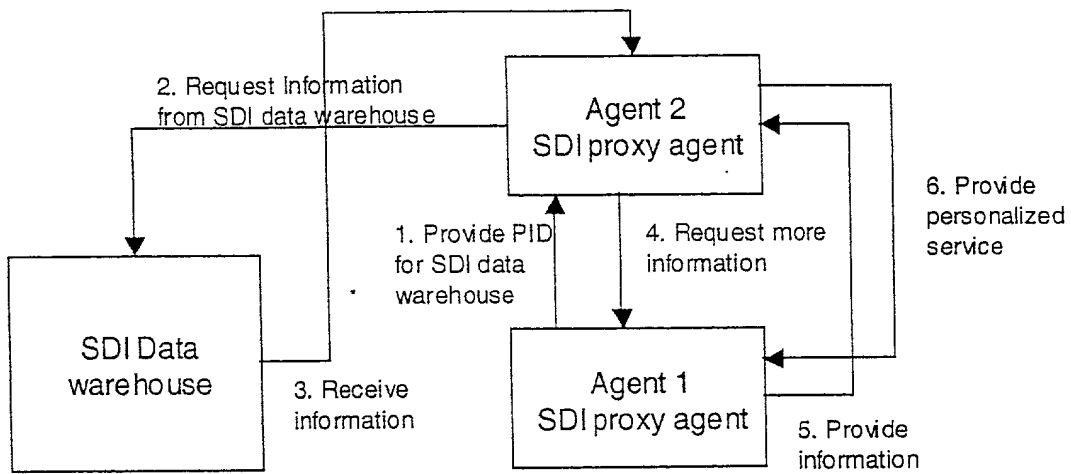


Figure 6: Data Stored Client Side and in the SDI data warehouse

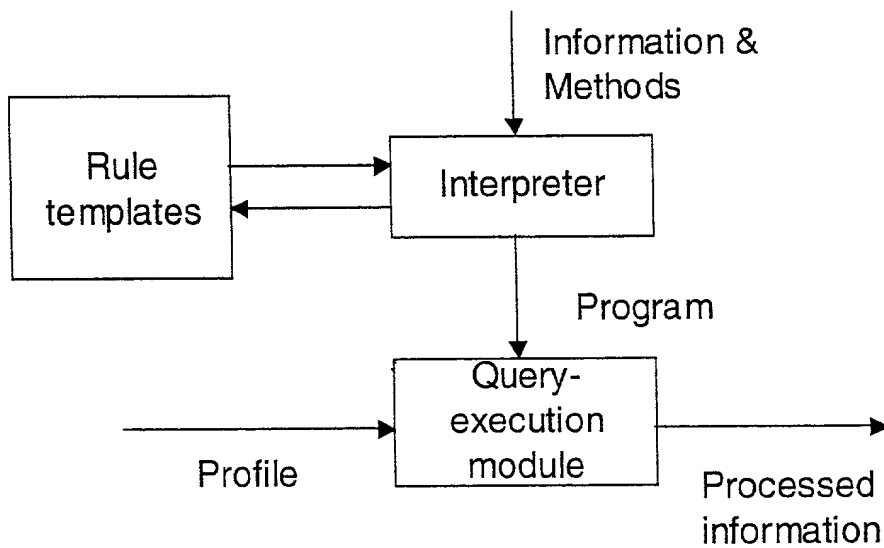


Figure 7. Client-side Personalization.

002207" 86066960

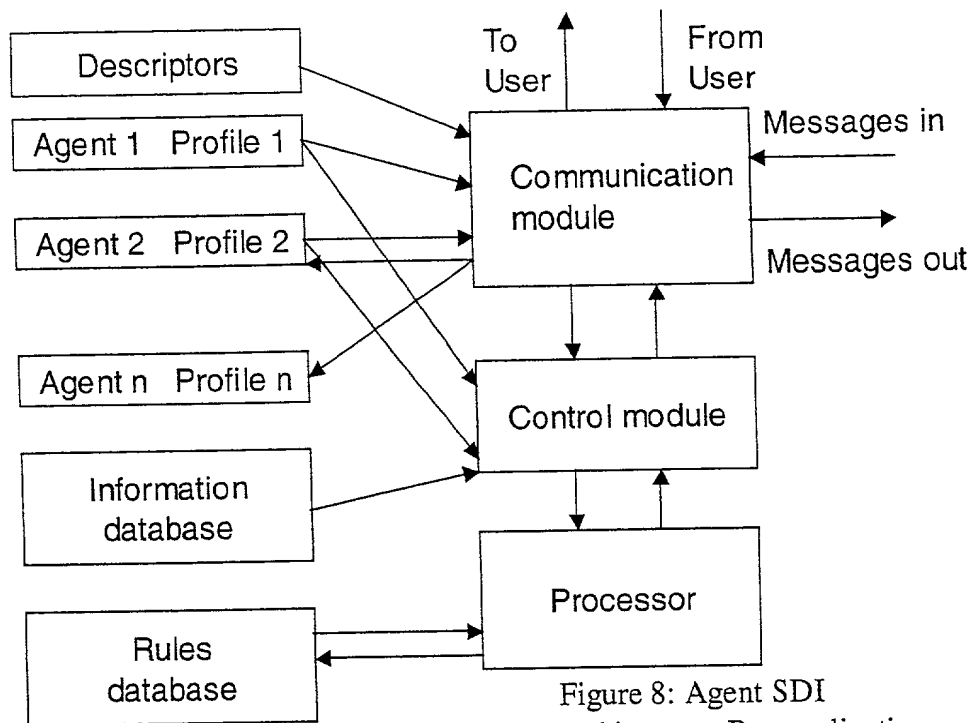
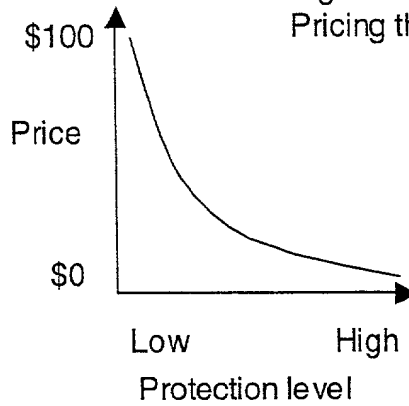


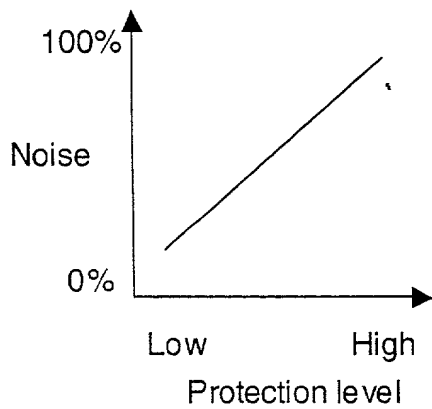
Figure 8: Agent SDI architecture: Personalization agent

002207" 86066960

Figure 13: Protecting Privacy:
Pricing the value of information.



Few Data records



Many Data records

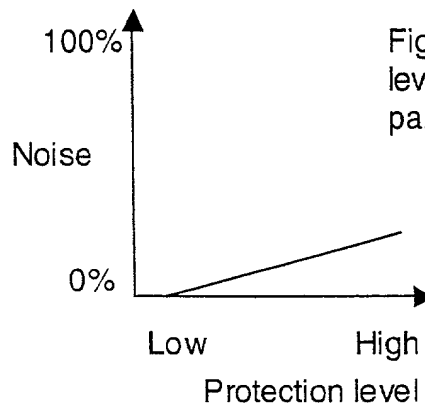


Figure 14. Selecting a
level of noise to provide a
particular protection level.

002207-86066960

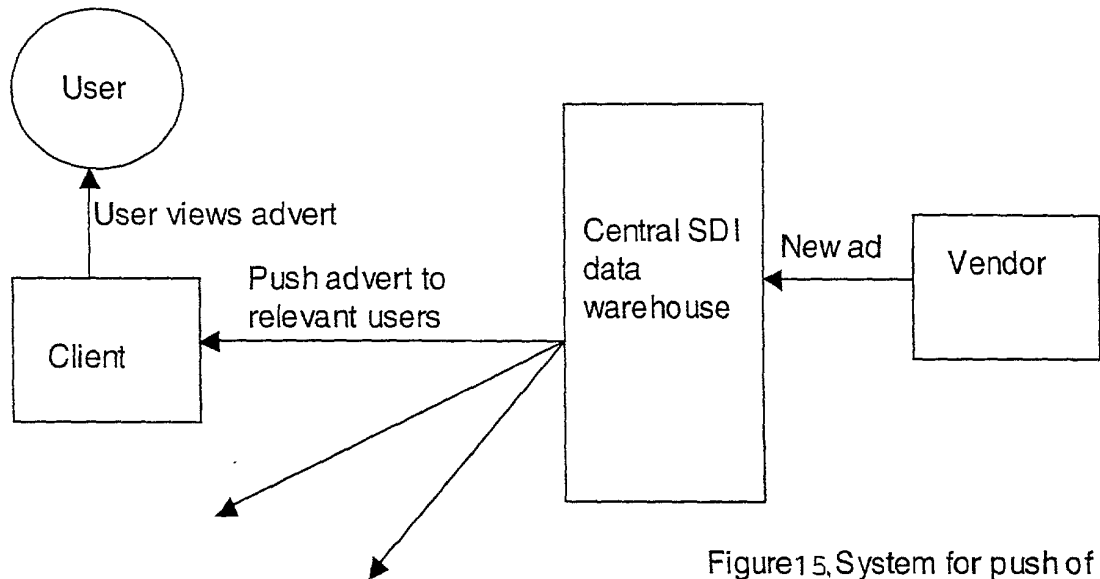


Figure 15. System for push of targeted ads to users in ianworthit.

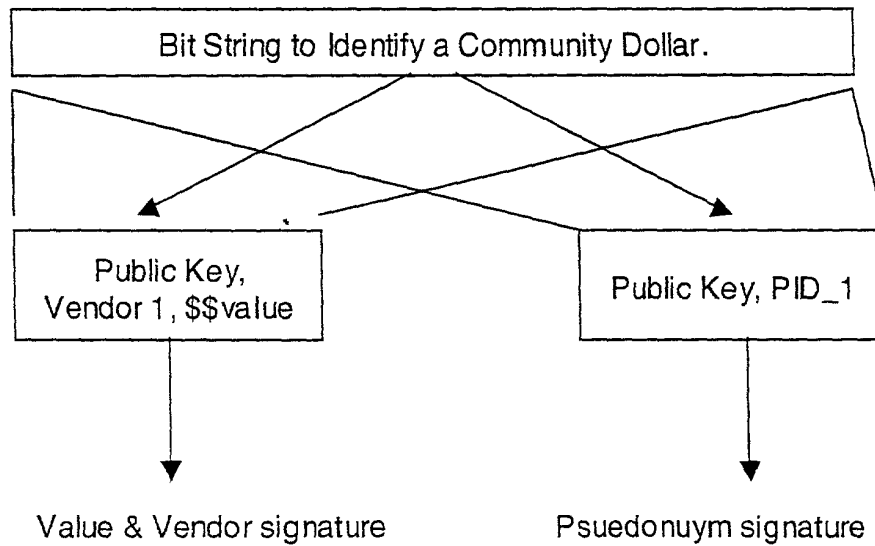


Figure 16. E-cash method to deliver community dollars.

002207-8605360

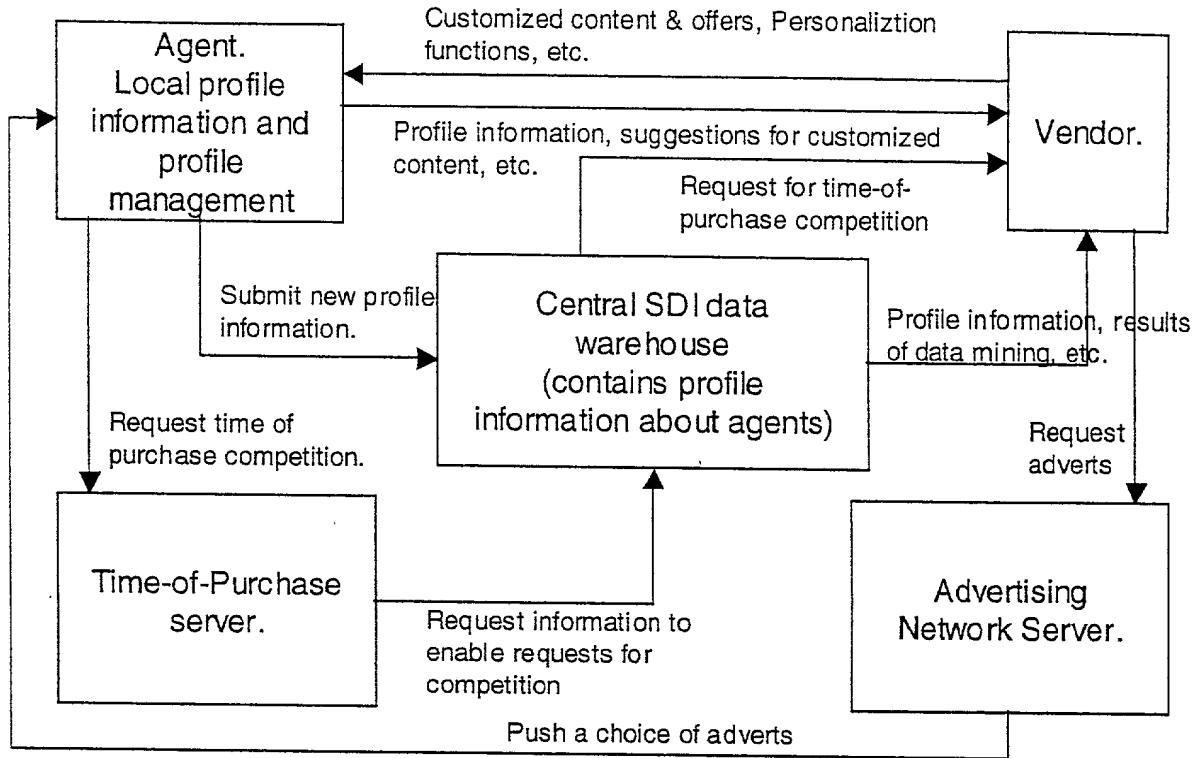


Figure1 7, Time-of-Purchase/lamworthit system

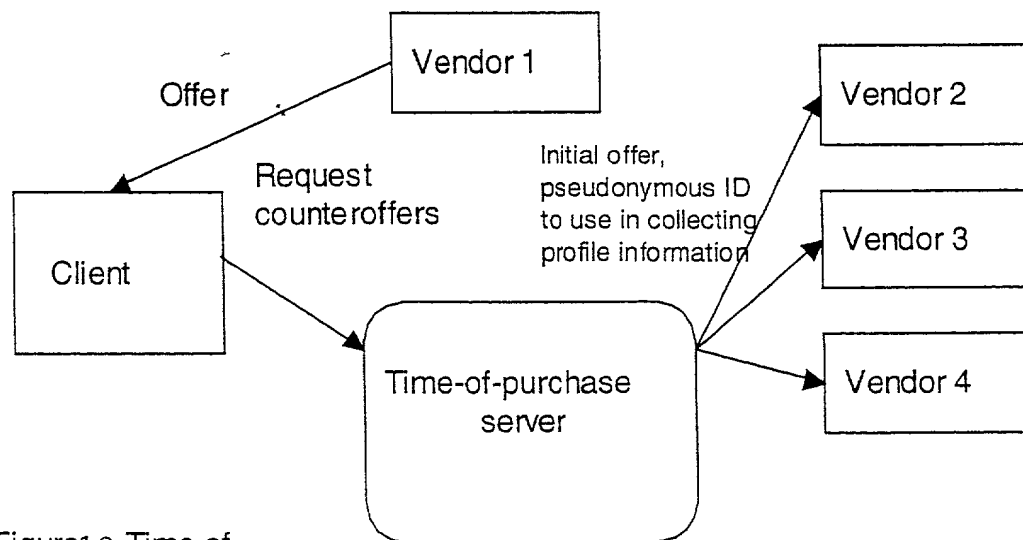
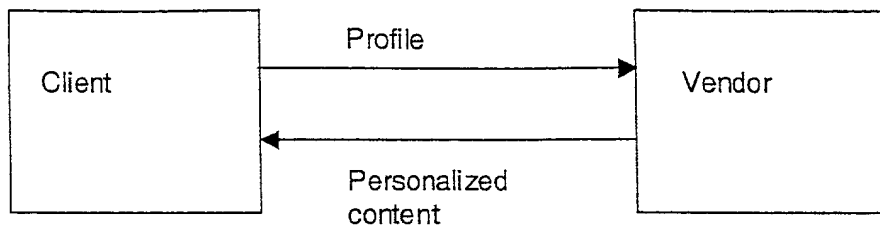
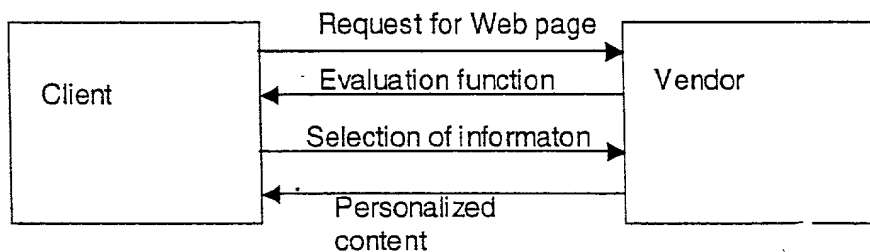


Figure1 8, Time of
purchase competition

0635MH-40874-103200



(a) Traditional Method



(b) Iamworthit/SDI Method

Figure 19.
Customization of
information to users.

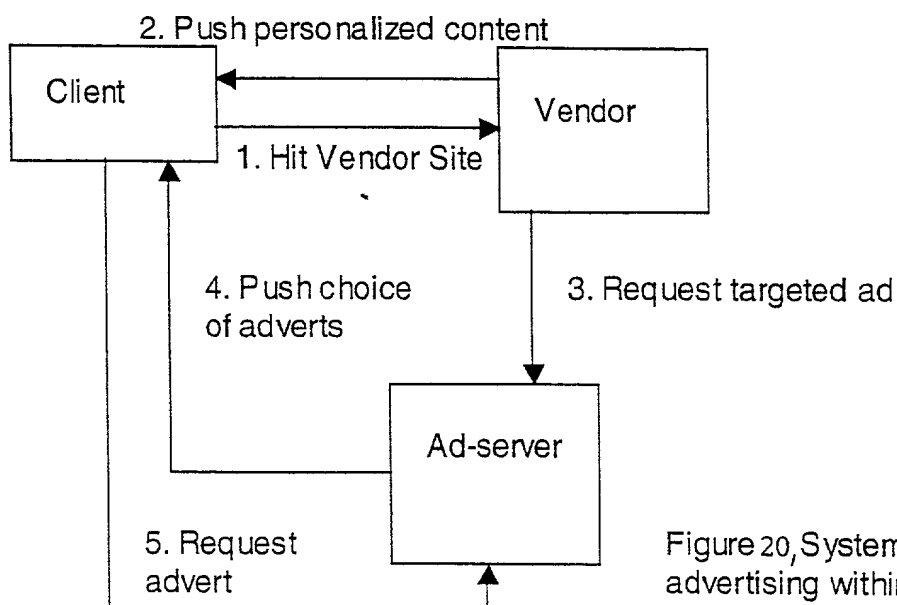


Figure 20, System for targeted
advertising within iamworthit.

002207-36066960

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Attorney Docket No. 0635MH-40874

In re Application of:

FREDERICK S.M. HERZ, ET AL.

Serial No. **TO BE ASSIGNED**

Filed: **HEREWITH**

For: **SECURE DATA INTERCHANGE**

§
§
§
§
§
§
§
§

POWER OF ATTORNEY

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

HERZ TECHNOLOGIES INCORPORATED, assignee of the entire right, title, and interest in the above-identified application filed herewith and being further identified by Attorney Docket No. 0635MH-40860, hereby appoints the following attorneys to prosecute this application and transact all business in the U.S. Patent and Trademark Office connected therewith:

Melvin A. Hunn, Reg. No. 32,574; **Kenneth C. Hill**, Reg. No. 29,650; and
James E. Walton, Reg. No. 47,245.

Send all correspondence to:

Melvin A. Hunn
HILL & HUNN, LLP
201 Main Street, Suite 1440
Fort Worth, Texas 76102
(817) 332-2113 voice
(817) 332-2114 facsimile
melhunn@hillandhunn.com (e-mail)

002207" 86056960

I hereby declare that all statements made of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or document or any patent resulting therefrom.

HERZ TECHNOLOGIES INCORPORATED



By: Melvin A. Hunn, Director

27 OCT 2000

Date

09690981.103700